UNIVERSITY *of*
TASMANIA
AUSTRALIA

MiniNauts: A Serious Game Platform for Enhanced Instruction in Reading Comprehension for Primary School Students

By

Aran James Cauchi-Saunders

(BComp. Hons)

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Tasmania

November 2017

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Aran Cauchi-Saunders

Abstract

This study presents the design, and results, from a serious game led intervention supporting reading comprehension in two Tasmanian classrooms. A series of mini-games were designed to support student's with in-class reading comprehension activities, with structured feedback opportunities presented between each game session. Learning content was enabled through an innovative Teacher Portal, where the game could be administered, and student progress could be monitored. An experimental design was utilised, with a pre- and post-testing methodology to determine if significant changes were affected over the course of an eight-week term. The results of this study found that all cohorts experienced a significant increase in reading comprehension scores between the pre- and post-tests; participants who interacted with a more gameful version of the serious game performed significantly better at reading comprehension skills when compared to their peers. It was also identified that feedback periods between game sessions were well received by participants but there were no significant differences between cohorts with respect to feedback mechanisms. Demographic factors were also explored, and the opinions of students were gathered and analysed with respect to reading comprehension achievement. This study concludes with a statement as to the viability of the tool in classrooms, along with manners in which it could be improved and upgraded in the future. Future research opportunities are also discussed.

Table of Contents

Contents

Table of Figures

Table of Tables

Acknowledgements

There are a number of people who need to be thanked directly, since without them, this Thesis would not exist. I am in no small debt to all of these people.

Firstly, I want to thank my effervescent supervisor Kristy de Salas, who somehow managed to keep me working, and was the generator of many of many of the best ideas in this project, I owe a very deep debt to her endeavours. I also want to thank Ian Lewis for his indomitably cheery disposition, and him providing me with many opportunities to grow personally and professionally for the many years now that I have known him. I also want to thank Andrew Fluck, for really believing in this project, and having a fresh perspective on many aspects of my research. Finally, to Lynda McKee and Natasha Visser, for being the wonderful and patient administrators of MiniNauts in their classes.

I would also like to personally thank everyone in the Games Research Group, with whom I have had many adventures—trials and adversities to be sure—but also many triumphs and moments of creation. I thank Lindsay Wells for being a much better programmer than me, and also for inviting me to what may have been the best wedding I have ever been to. I want to thank Adam Zulkifly for his ability to care deeply about things that had nothing to do with him, and for being a never-ending source of fantastic ideas and enthusiasm. To my 'Big Boy' Ben Geelan, I want to thank for being a really true friend, a sharer in many adventures, and source material for many of my insane tirades; I was only mad because you were usually right. Finally, Anthony 'Majestic' Smith, one half of Two Brothers, enough said.

Personally, I want to thank the special people in my life who shared my PhD journey, some of whom were there only for parts of it, but who built me up and made me a better person through their love, patience, input, and generosity. You know who you are ☺.

"We don't learn from experience. We learn from reflecting on experience."

– John Dewey

1 Introduction

This Introduction seeks to provide a brief insight into the research covered within this thesis, including an overview of the research background, objectives, methods and results. Further details on each of these can be found within their respective chapters in this thesis.

1.1 Background

Literacy is a fundamental component of living in a modern society (Campbell et al. 1996). However, there are many children and adults alike who suffer from illiteracy or have difficulty in comprehending some of the things they read (Maughan et al. 2009). This factor is made even more apparent in rural areas, where many adults are considered not even functionally literate (49%) (Australian Bureau of Statistics 2012). Research has pointed toward intervening at as young an age as possible when it comes to remedying deficits in student reading (Edwards et al. 2013). Demographic factors may play a role in problems with reading comprehension (understanding what one reads) when compared to their peers, it has been identified that motivating students to engage with reading is a key area where remediation in this deficit may begin (Willenberg 2005).

The question on how to engage students with respect to traditionally 'unstimulating' methods of literacy education is a large and varied field, with many proponents supporting a wide variety of educational and/or remedial solutions (Willenberg 2005). A solution that is becoming more popular over time is that of the technologically mediated solution, and in a format that students may enjoy more than others: video games (Gee 2005). These video games, termed 'serious games' aim to strike a balance between educational content and enjoyment for participants. Serious games have an explicit educational purpose and are not designed primarily for the amusement of the player. Another way to frame serious games, is as defined by Kapp (2012): *"an experience designed using game mechanics and game thinking to educate individuals in a specific content domain."* It too has been argued by Hunter (2013) that a new generation of

students prefer experiences that are visual, hands-on, repeatable, and most importantly digital, to match the growing trend of technology in their lives.

The field of serious games is dynamic and growing, research already exists around concepts to encourage player engagement, and educational benefit (Gambrell 2011; Kiili & Lainema 2008; Garris et al. 2002). Engagement of students with serious games has received much positive attention, and factors such as the flow state, whereby a player loses track of time and engages with the game for longer than they might expect; this flow state is considered a key component of successful video games, let alone serious games (Murphy 2011). Other factors such as intrinsic motivation in serious games may be a factor in compelling students to engage with these systems for longer and with more mental focus (Pavlas 2010), along with the feeling self-efficacy demonstrated by students with respect to their learning (Ryan et al. 2006).

Another key concept in the design of serious games is that of feedback and the role that reflecting on feedback has on learning outcomes, as well as for playing video games in general (Abrams & Gerber 2013; Zebel et al. 2013). Feedback has been demonstrated as a core component in learning, and the manner in which it is delivered to the individual learner is an important factor in the success of a learning attempt (Biggs & Tang 2011). Furthermore, serious games have been found helpful in providing opportunity for reflection during learning experiences, which Mezirow (1990) demonstrated to be valuable for learning and synthesising new knowledge. The manner in which to elicit these states in students is unfortunately an inexact science. However, new methods for eliciting behaviour change in individuals, have been described by behaviourists such as Michie et al. (2011), who presents a framework for designing interventions to remediate problem behaviours, called the Behaviour Change Wheel (BCW), and focus on desired behaviours. As previously identified, motivation is a core factor in engaging learners, and frameworks such as the BCW attempt to provide a starting point for remedial action for individuals.

The design of video games for education, or serious games, is a vibrant and fast-moving field. A subfield that has not been put to the test thoroughly, but which has been theoretically explored is that of mini-games (Isabel et al. 2014; Illanas Vila et al. 2008; Panagiotakopoulos 2011). Mini-games provide a small, digestible, and easily

designed/developed solution for creating serious games. Mini-games are cheap to design and produce (De Jans et al. 2017), and perhaps more importantly, it is easy for stakeholders (including students) to be involved in their rapid prototyping and development (Illanas Vila et al. 2008; Frazer et al. 2007a; Devisch et al. 2018; Frazer et al. 2007b). Series of mini-games combined together to form a longer conceptual serious game has been theoretically identified as a potential model for providing a platform for educational content, suitable for reading comprehension (Illanas Vila et al. 2008; Frazer et al. 2007b). The shorter structure of mini-games too allows a greater frequency of feedback to be cycled to the player, allowing them to reflect on their mental models of the learning content more often (Zapata-Rivera & Greer 2003; Cowley et al. 2013) The manner in which mini-games can enable learning, including literacy has been theorised to be practical, but no extant studies have focused on building a system of mini-games for the purpose of reading comprehension education. Factors of reducing the burden upon teachers to mark student work, or administer electronic worksheets tie well into the concept of automating game mechanics around educational content (Frazer et al. 2007a; Prensky 2008). More simply, the manner in which the burden of the game designer to produce educational content ahead of time for the classroom is absorbed by the teacher, who can customise lesson plans, and receive automated marking and reporting on student and class performance is an open, and as yet unexplored avenue. It is here that the opportunity presents itself to design a series of mini-games for the purpose of reading comprehension instruction. Mini-games present themselves as a vehicle for which techniques such as automating student feedback, encouraging more frequent periods of reflection, and more directly targeting problem behaviours through focused game design. Further details can be found in Chapter 2.

1.2 Aims and Objectives

The identification of the above described opportunity led to the development of the following research question: *To what extent can reading comprehension skills in students be strengthened by employing formative feedback within a system of structured mini-game experiences, and can this system satisfy the needs of students and teachers.*

This question covers several lines of inquiry. Firstly, the primary research question (and subsequent sub-research questions) aim to determine if a serious game for reading comprehension can increase reading comprehension ability in primary school students, when designed through a behavioural framework. Secondly, the research question seeks to answer how the factor of feedback, and the periods and depths to which it is administered is received by students, both in terms of measurable reading comprehension scores, as well as factors of motivation and enjoyment. Finally, demographic factors, along with the desire to quantify the success of the intervention with respect to teacher and student enjoyment and motivation need to be explored.

1.3 Overview of Methodology and Design

The overall research methodology of this study is described in Chapter 2 and Chapter 3, this section summarises the core methodology and system design of this thesis.

The physical output of this study is a serious game system, which came to be called MiniNauts. Design began through a process of requirements elicitation with relevant stakeholders in the education sector, including the inputs of two teachers, and their classes of students. This input informed a series of requirements for the software, including the identified undesired behaviours of students with respect to reading comprehension, and the desired outcomes of this study in terms of behaviours. Generally, teachers wished to see students be more motivated in their reading tasks at school, and to see reading comprehension scores increase. With this basis, the Behaviour Change Wheel was employed to design a set of required Behaviour Change Techniques and intervention options that would enable this behaviour change. The result of this process informed the next phase of the system design, which included mapping learning processes to game mechanics, through which the identified behaviour artefacts could be enacted using the Learning Model-Gaming Model (LM-GM) (Lim et al. 2016).

With the research aiming to use mini-games, as part of a broader reading comprehension package, a design was laid out for a serious game system called MiniNauts, that would include several mini-games, with game agnostic content whereby teachers could administer reading comprehension activities for their respective

classes, and have the game generate gameplay from these activities. Generated, and teacher curated feedback is presented to the player between the games and is intended to encourage reflection on player actions closer to the event relevant to the feedback.

MiniNauts was deployed to two classrooms for a period of eight weeks, with participants interacting with the software being divided into three cohorts (n=54). Firstly, players who acted as the control, and did not receive a gamified version of their reading content. Secondly, a group with gamified reading content, but without feedback mechanisms relaying information about their play performance. Finally, a third group was devised whereby participants would have full access to the game, as well as formative feedback relating to their reading and game performance read back to them at the end of every game session.

Quantitative and qualitative data was collected to answer the research question, with respect to reading comprehension performance, player satisfaction and motivation, the role in which feedback mediated greater test scores, and whether any demographic factors were an influence on player performance.

Details relating to the methodology and design of the system can be found in Chapter 2, 3, and 4.

1.4 Overview of Main Conclusions

This research generated several important conclusions for the serious game body of knowledge. Firstly, this research demonstrated that a novel connection between the BCW and game design frameworks such as the LM-GM was a viable methodology for designing a serious game for reading comprehension. A behavioural grounding has been previously identified as the possible starting point for serious game research, but as such, no extant research has demonstrated a clear progression from behavioural principles through a framework, toward a concrete game design. The practicality of this system was demonstrated by the overall satisfaction of the teacher and student participants, along with proving a stable, and modifiable platform for which serious games could embedded into a classroom and be administered by teachers. The discourse on mini-games as a viable platform for education, more specifically that of literacy education has also been broadened by this research. It was demonstrated that the system

was practical for digitising reading comprehension activities with minimal effort by a teacher, and having the games update responsively to this change in content.

This research also demonstrated a significant relationship between gamified reading comprehension activities, and a measurable increase in reading comprehension scores, when compared to a traditional reading comprehension activity, a change in mean reading scores of 18.63% and 23.16% for the two gamified test groups. This significant difference was demonstrated to not owe to confounding factors and is attributable to the gamification of reading comprehension activities that two of the three cohorts were exposed to. This positive result moves the body of knowledge toward a greater understanding of how reading comprehension can be successfully digitised and outperform traditional class-based reading comprehension activities. This opens the door for future research to explore game design for reading comprehension further.

The reading comprehension skills of players did not appear to be significantly impacted by the presence of in-game feedback mechanisms, which were hypothesised to improve player engagement, and reading comprehension scores. The automation of the feedback, whilst novel, did not provide enough of support framework to boost the reading comprehension skills of the students who interacted with it across the intervention. Students, while not significantly affected by the presence of feedback did however remain in a flow state while playing the game, as reported by a post-test survey. Whilst feedback was not a contributing factor, the options for digitising and automating feedback mechanisms have been explored by this research, and the balance between teacher setup, automation, speed of feedback delivery, and the meaningfulness of that feedback is still open for research.

1.5 Limitations

There are several limitations identified for this study. Firstly, due to ethical and scope considerations, a cohort of 54 students were identified and enrolled as participants in this intervention. As such, the overall generalisability of this study is low, however, effect sizes show that for higher sampling rates, a generalisable result may be promising.

Beyond cohort size, the design of the serious game included only two mini-games as part of the wider meta-game. For a longer study, the use of more mini-games

contained within a mini-game system could focus further on more specific reading comprehension techniques.

The experimental phase of this study ran for eight weeks across a semester in one Tasmanian primary school. The long-term benefits of this study were not measured through the use of follow-up surveys or testing, and the manner in which the benefits of the system manifest themselves in the classroom, months or years later is as yet unknown. Future research should focus on longer experimental periods and include an investigation into how the serious game is received by students and teachers in the mid to long term with regard to performance and user satisfaction.

1.6 Thesis Overview

The following chapters of this thesis will detail the rationale, design, game intervention, and results of this intervention. The chapter structure for this thesis is as follows:

Literature Review

- The role of serious games, with respect to literacy will be explored, concluding with a problem statement relating to gaps in the literature that present an opportunity for this study.

Methodology and Tools

- Leading from the problem statement of the Literature Review, this chapter will present the manner in which a framework for designing a reading comprehension intervention is utilised. Details relating to the experimental design are explored.

System Design

- The method in which the serious game was designed is explored in this chapter, relating back to theory and best practice.

Experiments

- The outline of the specific test procedure, along with details on cohorts, and participants.

Results

- This chapter will present the results of the serious game intervention, along with statistical analysis of relevant data.

Discussion

- Meaningful conclusions and comparisons are drawn in this chapter, relating to the statistical analysis presented in the preceding chapter.

Contributions, Future Work, and Conclusion

- This chapter concludes the thesis, drawing together the results of the intervention, presenting a conclusion to this work, while also expanding future opportunities identified as continuation points from this research.

2 Literature Review

2.1 *Literacy*

Literacy can be described as the ability to utilise, and the capacity to understand, abstract tools of expression such as language, art, and mathematics to describe and communicate with others (Campbell et al. 1996). Literacy can be broken down into forms of communication, and further into components of those systems (ACARA 2017b). Figure 2-1 visually describes the inter-related nature of these components.

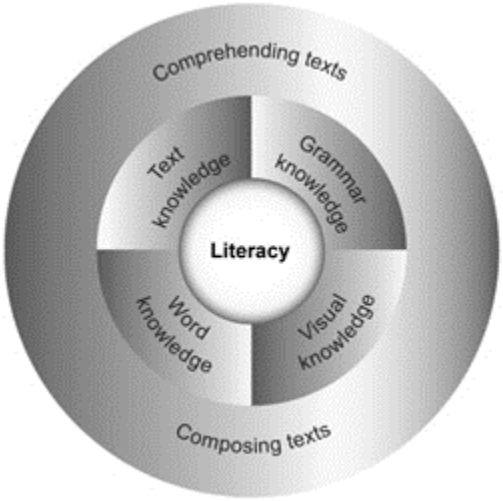


Figure 2-1 Components of Literacy Diagram (ACARA 2017d)

For this study, literacy refers to communicating with language, as opposed to other forms of communication such as artistic or technological. To narrow the focus of literacy even further, this study will regard literacy as components of reading and writing. The importance of narrowing the focus of literacy to the components of reading and writing, are discussed in the following section.

2.1.1 *Being Literate*

Campbell et al. (1996) describe literacy as a continuum of different, but inter-related skills, which are employed in a variety of ways across social contexts, as demonstrated in Figure 2-1. This definition of being literate can be seen as a more holistic interpretation of literacy, when compared to older methods of defining literacy as the ability to sign one's name, or read a basic form (Perfetti & Adlof 2012). Fundamentally,

literacy can be seen as the foundation to which other life-long learning is applied; and yet several cohorts of people are progressively becoming more marginalised, as a result of their illiteracy (Maughan et al. 2009). For children, the isolation that may arise from below average literacy can bleed into and negatively affect areas such as employment, societal obligations like taxes and documentation, being aware of complex community and world events and the education of their children (Cree et al. 2012; Australian Bureau of Statistics 2012). Machet (2002) argues that certain factors of early socialisation toward reading and writing can affect children's development of phonological and comprehension skills. These skills are critical to reading and understanding text, along with the ability to decode the words into understandable phonological processes (MacArthur et al. 2001). Without the ability to comprehend written text, part of a person's ability to engage in regular daily tasks becomes diminished (Richmond et al. 2012).

Literacy skills can be affected by environments and factors beyond the classroom; these factors include the child's race, access to healthcare and current health state, and their family's socio-economic status (Richmond et al. 2012). According to Cree et al. (2012), the individual's ability to catch-up to the required level of literacy as time progresses diminishes due to loss of motivation and social isolation. As the literacy gap widens, the individual may encounter fewer remedial services and fall into a cycle of not seeking help due to societal stigma which may compound the issue further (Roueche & Roueche 1999). It is therefore essential that children access literacy education which motivates them to learn, encourages their progress, and adapts to their specific learning needs.

Due to the inter-related nature of childhood to adult literacy, it is important to examine the effect of adult illiteracy, and as such, the ability to quantify the societal and personal costs of illiteracy in the adult wold. To relay the importance of pro-active remedial action in the adolescent world cannot therefore be understated; issues involving an adult's literacy stem directly from that individual's education as a child (Richmond et al. 2012). It is proposed that the earlier children are engaged with literacy, and maintain a minimum level throughout their schooling, the more likely that child is to experience normal literacy levels in their adult life.

Issues of sustainability for long term remedial action are a factor; some remedial programs may not be adequate for creating large changes over long periods, but small and short-term programs—or interventions—that aim to improve a child's literacy, motivation and scholastic satisfaction may be of some use (Willenberg 2005). While structural reform at a macro-level is necessary to ensure that every child receives equitable opportunity (Richmond et al. 2012), micro-level changes can be made to the classroom simultaneously, that directly affect the child's perception and performance in literacy based tasks (McCutchen et al. 2002).

2.1.2 *State of Childhood Literacy in Australia*

According to Australian Bureau of Statistics (2012) Australia lies in the middle of the spectrum when compared to other OECD countries with respect to observing overall literacy levels in the adult and youth population. Literacy in the context of a study by the ABS (2012) includes the domains of prose, numeracy, document and problem-solving. Respectively, the four literacies are concerned with: comprehension of documents, ability of the individual to apply mathematics in a variety of situations, the ability to utilise information resources to accomplish tasks, and ability to generate own solutions to unique problems. The ABS define a five-level ranking system to describe where an individual's literacy competency resides, Table 1 details this further.

Competency Level	**Description**
Level 1	Little to no literacy skills
Level 2	Below minimum, struggle day-to-day
Level 3	Minimum to meet societal obligations
Level 4	Above minimum, competent
Level 5	Highly competent literacy skills

Table 1 ABS Literacy Rankings (Australian Bureau of Statistics 2012)

Concerningly, in 2012 it was recorded that approximately 18% of people in Australia achieve a Level 1 literacy rating, which means that the individual suffers from a low to very-low level of basic literacy (Australian Bureau of Statistics 2010). When comparing these statistics regionally, Tasmania (a regional, low population density State

of Australia) has a population in which 20% of people achieve Level 1 prose literacy, and 21.4% attain Level 2. 29% of people achieve Level 1 document literacy, while 29.3% of people achieve Level 2 for document literacy (Australian Bureau of Statistics 2010). Cumulatively, this means that Tasmania has a 49% rate of people achieving below basic literacy levels for prose literacy and a 50.7% below basic literacy rating for document literacy (Australian Bureau of Statistics 2012); with such rankings, Tasmania can be considered to the lowest performing State in Australia with respect to literacy. This lack of student attainment regarding literacy skills is a worrying statistic, made even more dire by the well-linked outcomes of lower-attaining students who do not have functional literacy skills to problems of overall wellbeing. The cost to individual health and wellbeing, has been documented, with links between poverty, poor health and poor job outlooks for people with low literacy (Australian Bureau of Statistics 2012). Along with the societal costs of illiteracy, estimated productivity loss due to the low literacy standard in Australia costs $18.35b in lost GDP (Cree et al. 2012). It is again important at this point to reiterate that adult literacy problems are highly connected to childhood education (Perfetti & Marron 1995). Specifically, in Australia, the number of students who attain the national average of reading competency in year 5 is 92.9%. In Tasmania, this figure drops to 91.3% (ACARA 2014). When considering gender disparity, male Tasmanian students consistently fall behind female students across the national and state averages. Boys are over twice as likely to be below the national average for reading level at 9.0% in 2014 as for girls at 4.9% (ACARA 2014), around age 11. While less than one in ten students suffers from very poor literacy in Tasmania, the literacy gaps dramatically across age categories, trending toward the 50% average adult illiteracy level presented (Australian Bureau of Statistics 2012).

As presented so far, the ramifications of illiteracy reach from childhood into a person's adult life (Perfetti & Marron 1995). Access to employment, along with higher risks of poor health may be affected by deficits in a person's childhood literacy levels (Rivera-Batiz 1992). It has also been shown that illiteracy is a serious problem for Australian—particularly Tasmanian—students and that students are regularly falling below the national standard for functional literacy. This problem is exacerbated in males, who can be seen as at higher risk of developing literacy problems by Year 5

(Australian Bureau of Statistics 2010). The intent of literacy education is in part to attract student's attention toward learning; with a child's attention and engagement on literacy, the later-life ramifications of poor literacy may be averted (Marsh & Hallet 2008). To successfully address the issues with literacy, one must first examine where the barriers lie that inhibit successful literacy outcomes

2.1.3 *Barriers to Effective Literacy Education*

It has been identified in the preceding sections where literacy issues manifest themselves in the lives of children and adults. Contextualising how and why these issues arise in children, particularly those identified as most at risk—males, in predominantly regional states such as Tasmania—is therefore a key consideration.

Snow & Matthews (2016) contend that while certain reading programs, remedial or otherwise, may have positive effects on particular cohorts of students, these benefits can sometimes be constrained to students from higher socio-economic backgrounds; it is perhaps the students who most need the help do not receive it as effectively as others (Snow & Matthews 2016). The manner in which these teaching programs are delivered by the teacher also has a role to play in the state of literacy for students; often, the language and expression that teachers employ in the classroom can have an effect on student literacy levels. Teachers may have access to well-regarded literacy programs, but due to funding, or professional development, are unable to deliver these programs efficiently or effectively. The teacher in this sense, is seen as the gatekeeper to student literacy acquisition, and student performance is linked to their teaching methods. It is therefore necessary for teachers to take a pro-active role in examining new techniques for lagging students, and identify trouble areas where students may be falling behind with traditional methods (Ertmer et al. 2012). Teacher beliefs have been identified as a prime factor in how teachers approach new methods of teaching; a teacher without a student-first focus has been shown to deliver poorer outcomes for students when new delivery methods were prescribed. Further, Williams (2014) state that another barrier to students successfully acquiring competent reading skills is the lack of opportunities for students to perform sustained reading opportunities in the classroom, such as silent reading during class time. Students should also be active in the selection of reading

materials, but many teachers do not have the time or resources to enable this policy to take fruition in their classroom (Williams 2014).

While the issues presented are structural in nature, other problems at the student level may affect how well children acquire literacy skills over their childhood education, for example: young males consistently score lower in literacy scores, particularly in Australian schools (Australian Bureau of Statistics 2012). More generally, issues relating to interest and motivation with respect to the method, and content of literacy materials has been identified. Students, according to Topping (2015) consistently read at higher levels when reading fiction, over non-fiction. Ivey & Broaddus (2001) contend that students when given no choice over what they read, will experience significantly lower motivation to engage with the material, and get meaningful outcomes from the activity. According to Henry et al. (2012), boys see education as significantly less important than girls, and as such, don't start at the same motivational base. To exacerbate this boys also may have to contend with stereotypes about the use and perception of reading over alternate activities (Henry et al. 2012).

While the issues presented form a barrier for many students to gain competent literacy skills in line with their peers—both in the classroom, and nationwide—research into systems to remediate these issues is ongoing. The following sections will explore these such solutions.

2.1.4 *Closing the Gap*

2.1.4.1 Back to School

Many countries encourage adult learning of basic literacy in an attempt to redress the problem of childhood difficulties in literacy (Purcell-Gates et al. 2002). While the effects of remedial action towards literacy in the latter years of a person's life may help to redress some of key literacy deficiencies, this is still far from ideal. Ensuring that illiteracy does not create disparity at a young age is of key importance; the age at which learners are encouraged into remedial literacy programs is a direct contributor to the success of their remediation (Edwards et al. 2013). To capture learners who may be falling behind, and effectively provide remedial literacy opportunities, the need to keep students engaged in school is a key consideration (McMillan & Marks 2003). Increasing

the retention rates of students in schools is another key consideration, whereby students are not legally allowed to leave school until a mandatory age has been met, rather than allowing students with poor literacy to remove themselves from potential solutions (McMillan & Marks 2003). Keeping students in school is only part of the solution; the need to effectively spend time instructing students in a manner which is most relevant, interesting, and beneficial to them is apparent (Applegate & Applegate 2010).

2.1.4.2 Whole Language Learning

A popular solution to the literacy gap, according to Lê & Lê (2007) is the drive to return to the basic education pattern of teaching grammar, spelling, and vocabulary supported by regular testing, in a holistic manner. This method relies heavily on paper-based, prescribed activities that students work through at the pace it is taught by the teacher (Lê & Lê 2007; Lê & Lê 2007). Students are expected to perform literacy activities as composite activities, whereby a complete meaning of the text is generated; more simply, these activities encourage multi modal instruction, with an emphasis on reading what is written, re-writing, and discussing. In the words of Weaver (1988) *"Anything I can say, I can write; anything I can write, I can read"*. While this method is a popular in Australia, termed the 'Whole Language Approach', it has been criticized for its lack of focus on catching issues are they arise with individuals (Coltheart & Prior 2006). Part of the issue with not quickly remediating problems in individual's literacy are the limitations of the teacher's time spent per student (Ivey & Broaddus 2001). Spending time with students, correcting mistakes and providing instruction, is considered a vital component of literacy education (McCutchen et al. 2002). However, direct access to the teacher is a scarce resource, and students may only have limited one-on-one access to the teacher to mark and correct their work. It is well established that regular correction and encouragement by a teacher can improve learning outcomes in students (Trigwell et al. 1999). While systems for encouragement may not be easily automated, systems to assess student work and provide feedback can and have been automated, most typically through technology (Lê & Lê 2007).

2.1.4.3 Technologically Mediated Solutions

Solutions to the literacy problem may include technological mediation. Literacy, according to Dede (2005) needs to be redefined to include the increasing relevance of the internet and digital communication in classrooms. As Prensky (2001) contends, children are becoming increasingly 'native' in the world of technology; with greater immersion of technology permeating their cultural and social connectedness. It is thus the role of educators to facilitate education to modern children in a form that most accurately reflects their world view and situation in the opinion of Prensky (2001). Students who consume technology in their personal lives are however, suffering a disconnect, due to the technology in their school rarely aligning with what they are familiar with (Dede 2005). Therefore, students' literacy needs may be better served by educational technology which matches their expectations of technology (Lê & Lê 2007).

Games such as *Mathletics* (Learning 2017) attempt to provide a technologically mediated solution to the problem of motivation students to engage with mathematics (Table 2). Students are presented with a wide variety of mathematical tasks and compete with students throughout Australia, their State, and the school with respect to score and reaction time. This service is highly popular in Australia, and has been used by hundreds of thousands of Australian primary school students over the last decade (Learning 2017). Whilst this service does not serve literacy content, the mechanisms in use are repeated for other literacy games. Services such as *Reading Eggs* (Blake eLearning 2017) and *Grammatikus* (Sharp & Fitz-Gerald 2016) attempt to provide technological solutions to the literacy problem (Table 2). The driving ideology behind these services is the assumption that children will respond to technological activities more favourably than to paper-based solutions. Presenting literacy objects with multimedia has been explored in great detail, with many studies suggesting that content displayed in a variety of audio and video formats can increase literacy levels compared to traditional activities (Tjus et al. 1998; Sorapure et al. 1998; McKenna 2006; Segers & Verhoeven 2002). These systems present literacy in a variety of technological formats but rely on similar principles of audio-visual accompaniment to basic literacy activities. The way in which the *actual* literacy object is presented is unchanged, with services such as *Grammatikus* relying upon simple question-response interactions.

Title/Genre	Approach	Limitations
Reading Eggs (Mini-game Exploratory)	• Allows content agnostic literacy lessons to be embedded • Easily driven by the learner • Many modes of teaching lessons present	• No flexibility in content manipulation • Requires expensive hardware for students (tablets)
Mathletics (Math Quiz)	• Fast, repeatable, and comparable question and answer format maths questions. • Players are ranked on an individual, school, and state basis for motivation factors	• Does not gamify its maths content. • Relies upon gamification techniques such as points, and leaderboards to motivate students.
Grammatikus (Role Playing Game)	• Create a character for customisation purposes to encourage connection with the player. • Level up system to encourage repeated play. • Strong feedback mechanisms, and report driven feedback for students and teachers	• Costly license • Requires teacher onboarding and training • Requires significant in-class commitment regarding time allocation • Content cannot be modified by the teacher for the needs of the classroom

Table 2 Serious game comparison

Ertmer et al. (2012) argues that teachers who are more open to technological change in the classroom, provided they are coming from a student-centred approach, will have greater impacts on student learning outcomes than teachers who use technology just a prescribed, or are slow to upgrade their teaching methods Ertmer et al. (2012). For technologically mediated progress to occur, teachers need the support of

both administrators and the wider community who believe that technology can aid students in their learning, not as a gimmick, but as a fundamental supplemental tool for engaging and teaching students. Watson & Yang (2016) found that teachers who had previously engaged with technologically mediated solutions in a classroom setting, were much more willing to try new techniques repeatedly and not forsake the entire concept entirely. This is a heartening discovery, as it allows researchers the opportunity to get the foot in the door with their educational software, and if it is proven effective, have a reliable champion for future endeavours and collaborations. With the barrier of teacher approval hypothetically conquered, it is time to draw attention to how technology can be designed in a way that supplements this enthusiasm and buy-in from educational stakeholders.

2.1.4.4 Play

A strategy for embedding literacy education into technologically mediated systems is through the use of gamification. Gamification, as defined by Dominguez et al. (2013) is the application of game-like components into non-game environments. Keeping score of a whole class' reading performance and awarding digital rewards to those who have completed the most reading tasks could be considered a gamified experience; the scoreboard models a form of competition, while the rewards attempt to engage goal seeking and motivating behaviours. Gamified systems have an inconclusive track record in generating long term and stable learning outcomes, according to Dominguez et al. (2013), and (Susi et al. 2007) but can generate positive effects dependant on the specific context (Hamari et al. 2014). Gamification, at the intersection of play, and behaviour change such as learning are what described as a points system seeking a purpose (Nicholson 2012). The long-term benefits, the casual nature in which gamefulness is integrated into the system, and the manner in which achievement in the system is prioritised over meaningful changes in behaviour is of further concern (Nicholson 2012). There are also allusions that gamification eschews traditional mechanisms for play, prioritising instead systems of points, and rankings as motivators themselves (Hamari et al. 2014).

As mentioned, the intersection of play, and behaviour change do not necessarily logically conclude at gamification. The manner in which playfulness, and gameplay is

integrated into the secondary goal of the system is still very much up to the designers; the ratio of play to time spent on other tasks can be adjusted (Kapp 2012).

Sutton-Smith (2009) assert that play is considered an important tool for children to learn and engage with the world. Many life, social and personal skills are learned through the act of play. Children and adolescents engage in significant amounts of play during their lives, either with toys, other children or their own imaginations (Marsh & Hallet 2008). It is hypothesized that play provides a significant opportunity for the accretion and practice of language and communication skills. Kervin (2016) argues that an increasing body of work in the space of play and literacy has opened the opportunity for new avenues of combining play, with new modes of delivering educational materials. Steinkuehler et al. (2010) contends that video games for learning *"sit in a complex and productive ecology"*, and that when tasks are *interest-driven*, that is, they intersect the player's interest on a topic with educational content, the student has an edge which does not exist for tasks that are not encompassed by their interests (Beavis 2017).Video games are a natural extension of both technology led interventions for literacy, and playful interaction for children (Charsky 2010), and their use as a medium for literacy education has a long history. It is the conclusion of Kervin (2016) that while technological solutions to the literacy gap are of great importance, current systems do not go far enough to engage students long-term, and in the technological language that children are familiar, that of video games (Ma et al. 2011). In the following section, video games for the purpose of learning are discussed, and the various mechanisms through which the secondary goal of literacy may be supported are examined.

2.2 *Video Games and Learning*

Garris & Driskell (2002) describe the umbrella term 'games' as a form of entertainment whereby a person can voluntarily engage in play with a system which provides no tangible product but induces behaviour of rule-following and enjoyment. A video game is the digital representation of such an activity. Video Games worldwide are responsible for a several billion dollar industry focused on creating video games and video game hardware for consumers (Marchand & Hennig-Thurau 2013). A long standing attraction between video games, children, and adults has allowed a plethora

of games spanning a wide variety of genres to emerge (Gee 2005). As video games seek to engage players in playful behaviour, a variety of mechanisms are designed (either implicitly or explicitly) to attract the player's attention for a significant amount of time (Lieberman 2010). Video Games are composed of a variety of systems and processes which describe the player's behaviour, thinking and motivation during and for play. The following sections will detail several components of video games which have been described as key components of video games.

2.2.1 *Video Game Mechanisms*

As previously stated, video games allow people to engage in entertainment which enables playful behaviour (Sutton-Smith 2009). Rieber (1996) argues that games fall into the growing trend of instructional technology and defines play as being voluntary and intrinsically motivating, engaging and may contain elements of imagination and make-believe. As with any entertainment media, the goal of video games is to engage the participant of the activity; specific to video games is the desire for users to engage in play or playful behaviour (Murphy 2011). To enable this playful behaviour, mechanisms which support the player in choosing to fully engage with the video game must be examined (Michael & Chen 2005). One such mechanism for better engaging the player's attention is that of the flow state, and the way it affects player attention and desire to continue with the video game; this will be discussed further in the following section.

2.2.1.1 Flow

Video games, according to Murphy (2011) enable the experience of a human performing a task that is highly enjoyable and engrossing; this is termed the flow state. Csikszentmihalyi (1990), contends that people are happiest during the state of flow; and that fundamentally, the difficulty of a task must be proportional to the skill of the person undertaking it, for the necessary conditions of flow to arise. During the flow state, a person's attention is maximally drawn to the task at hand, to the point that the person may disengage from a variety of other tasks to continue the flow task. The player of an immersive video game may choose to eschew eating, or observe that time has passed much quicker relative to their normal perception of it (Pavlas 2010). While this state may

appear unhealthy, it is actually the application of a person's undivided attention, and the powerful effects of this attention can be seen in the acquisition of skill that person has with respect to their flow-oriented task (Csikszentmihalyi 1990). Murphy (2011) asserts that the flow state is a balance between the efforts expended on the task against the player's skill level. More simply, the goal must be within reach of the player's abilities, without being too easy as to bore the player, listed in Figure 2-2. This Flow Zone accounts for a window where the person is receptive to the medium attracting their attention and are highly compliant in new challenges that are thrown at them. Murphy (2011) contends that the golden opportunity for influencing a person can occur during this window of maximal attention. Simply, the player has a great well of motivation at this point, and that motivation has the alluring promise of high-focused attention on a given topic or activity. Great care needs to be taken however to not tip the individual out of the flow state, and toward the anxiety or boredom axes.

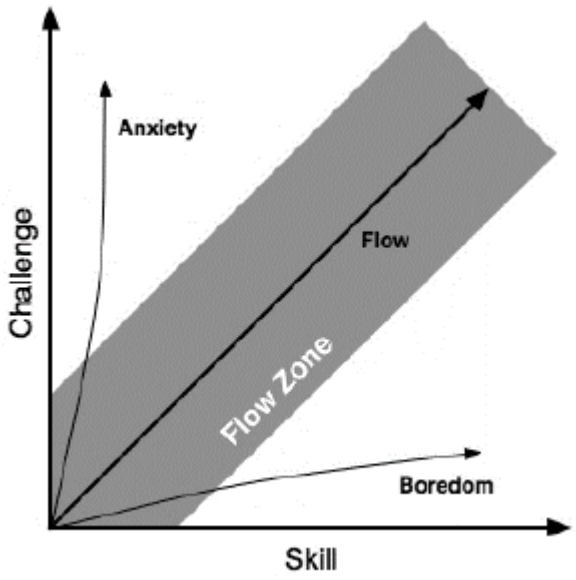


Figure 2-2 Skill-challenge relationship (Pavlas 2010)

Essentially, flow is a state of intrinsic motivation of the player, whose desire to engage with the game is motivated from their internal need to play, win or socialise in the game (Pavlas 2010). Kiili & Lainema (2008) describes playability, gamefulness and story as necessary antecedents to flow for video games. With respect to flow and learning, Kiili (2007) argue that the flow state must engage the player to induce their undivided attention, but not stand in the way of active and participatory engagement; simply, the

person should have the motivation to continue their activity via their own volition, and never be pressured to perform an action they do not like to the point of distraction (Garris et al. 2002). More simply, game players must enjoy the game, but also be allowed moments to actively reflect and receive feedback on their progress toward the secondary goal of learning; if a player were to simply play a game and then move on to another task, potential learning opportunities may be missed. This ties into the iterative effects of feedback cycles, discussed further in Section 2.2.1.3. The state of flow is best where cycles of struggle, and then achievement are made through a continual process of reward and opportunity (Csikszentmihalyi 1990). Where flow is induced, the desire to continue that state of flow, or return to it, is referred to as motivation, and is considered an integral component of video game systems (Marsh 2011).

2.2.1.2 Motivation

Motivation is a core component of the attraction of video games (Mekler et al. 2015); motivation too is a fundamental component of learning, or lack thereof (Garris & Driskell 2002). The junction between video game motivation and learning motivation is where a growing body of knowledge lies on the subject of effective and motivating educational games; this will be described in the following sections. Garris & Driskell (2002) asserts that motivating players to engage with games are six defining characteristics of video games based on a meta-analysis of literature: fantasy, rules/goals, stimulation, challenge, mastery and control. When video games successfully integrate these features; play, flow and enjoyment of the game are quick to follow (Pavlas 2010). Ryan et al. (2006) argue that self-efficacy demonstrated through autonomy and competence within the game are of primary important to the motivation to continue playing a video game. These states are where a person is primed for becoming involved in a task for longer, experiencing deeper involvement and engaging more deeply with the video game (Garris & Driskell 2002). What is demonstrated from the aforementioned game characteristics to induce motivation in video games, is the inherent desire for control, stimulation, and self-efficacy; the harnessing of these drivers to play are what Kervin (2016) describe as being an opening for digital technologies to bridge the motivational gap that exists for students in the literacy space.

2.2.1.3 Feedback

Interaction, according to Gee cited in (Kong et al. 2010), is where a system engages with a user in the form of feedback. Kong et al. (2010) describe a defining feature of a 'good' learning system as containing the concept and practical implementation of interaction. Gee cited in Kong et al. (2010) argues that interaction is vital to learning. Feedback in learning is the process of being advised as to the outcome of the practice of a learner's knowledge by an authoritative agent Hattie & Timperley (2007). This can be seen as the 'output' stage of the input, process and output model of learning and as described by Hattie & Timperley (2007, p.83) is *"the consequence of performance."* The learner enacts their knowledge in a quantifiable manner so that it can be validated by another agent (typically a teacher or trainer) so that the learner can be advised as to whether their knowledge of the content is considered complete or any level below that (Hattie & Timperley 2007). The initial instruction of a learner can be clearly separated from the final evaluation of the student, or the two can be deeply intertwined to produce what Sadler (1989) describes as closing the gap between what the learners understand and what they are attempting to learn. This concept directly relates how feedback can be used via video games to demonstrate achievement over time. Context clues could also possibly be integrated as feedback mechanism for serious games (Rupp et al. 2006); positioning reading material within a clear context can provide a grounding for greater understanding of the meaning of the sentence or the word's place in it (Kintsch 1994). Feedback can also be described as the educational principle of formative assessment according to Biggs & Tang (2011). Formative assessment is the process of monitoring the performance of a student through low-hanging methods of determining the state of a learner's knowledge; formative assessment therefore exists as a complementary measure to summative assessment, which aims to determine whether something has been learned (Biggs & Tang 2011). As described by Garrison & Ehringhaus (2007), formative assessment is considered a complementary, and indeed a base component of summative assessment. To form a summative assessment protocol, the teachers are in a position to digest the quantitative feedback data and produce feedback that is easily interpreted by the learner (Abrams et al. 2018). The purpose of this summative assessment is to bookend the student's learning, and to provide them with a reasonable

closing to their learning activity, with measure for how they are to continue in the future (Abrams et al. 2018).

Hattie (1999) performed a meta-analysis of several hundred studies on the effectiveness of feedback as an agent in student learning in a wide cohort of schools. The study found that a variety of feedback measures were of varying usefulness to students. The most important feedback measures affecting student performance were cues and reinforcement, along with visual, aural and computer based feedback. A more systematic approach to determining exact feedback measures was described by Kluger & DeNisi (1996), who described controlled experimental designs which control groups and attempted to reduce the number of confounding variables at play in the tests. It was found that in the general case, providing cues on correct performance for students was of greater importance to learning outcomes than promising tangible rewards or praising the effort of the individual. Hattie & Timperley (2007) describes a feedback model which attempts to reconcile the problem described previously of the 'gap' in a student's understanding. The three questions of feedback described in the model in Figure 2-3 attempt to describe the process a learner needs to undertake to fully grasp the gap in their knowledge to make their learning efforts more successful. 'Where am I going?' is the framing of a question as to the exactness of the goal being attained. Fully understanding the goal of the learning process is key to successfully engaging and then understanding what is required to learn the desired educational content. If the goal is not well communicated or the goal is unattainable via its complexity or the required effort to attain it is too great, the learner may disengage from the activity. The act of failure in this cycle of goal based iterative learning is also important; as noted by Anderson et al. (2018) failing allows the player space to re-evaluate their approach a given task, and given proper feedback, can use the discomfort of losing as a driver for greater motivation in subsequent cycles. The interplay of feedback with learning goals also extends to the previously mentioned criteria of well-formed feedback for students. If the goal is explicit, yet the feedback fails to address confirmation on correctness or cues as to the state of the current learner's knowledge, the benefit of feedback is being wasted. Goal setting does not fully encourage a learner to engage in a task, it is up to

other educational processes to perform many motivational forces for a student to learn successfully.

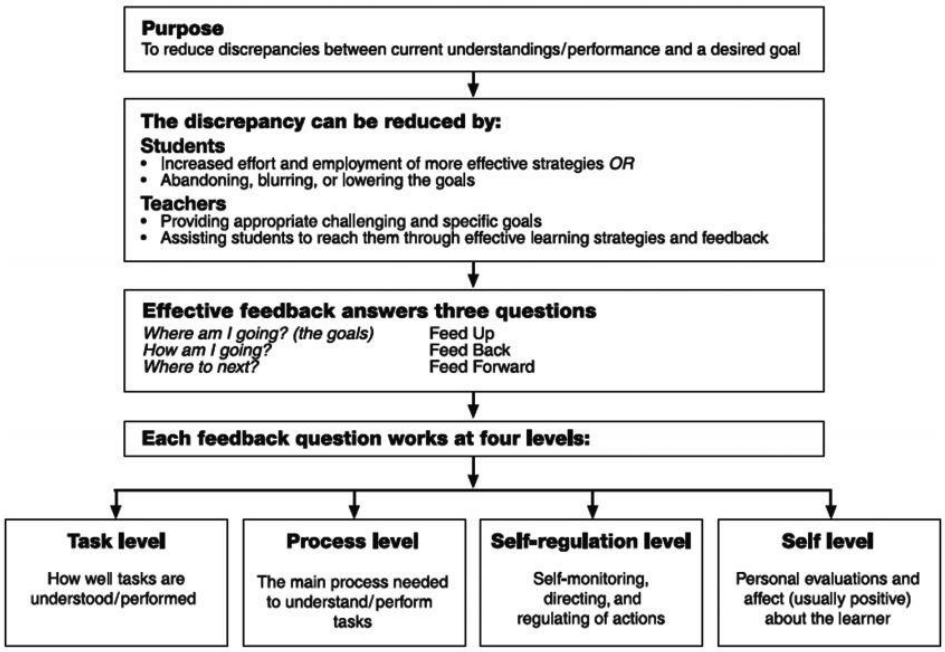


Figure 2-3 Hattie & Timperley (2007) Feedback Model

The second question according to the Hattie & Timperley (2007) Feedback Model is 'How am I going?' or more critically; how is my actual performance when compared to my expected performance? Importantly, this stage of the Feedback Model is where the role of feed*back* becomes essential via the conduit of a teacher. Primarily, the learner's knowledge is then in some way assessed—via a test, questionnaire or simple conversation—their performance is ranked based on a 'known' adequate level for a learner of that ability, age or competence. The teacher then may choose to provide feedback in a variety of forms to the learner ranging from none at all, to some of the previously mentioned factors such as focusing on correct answers, coaching or associating a correct outcome with an expected reward or behaviour.

The final step in the feedback process for a learner is structuring the question 'Where to Next?' Hattie & Timperley (2007) sates that this component of the feedback model is where the teacher may engage the learners' attention toward new goals, introspection on their performance, and strategies to aid them in the future. The role of future projection is important to the principle of feedback and learning in general. The

ability to look forward to new goals, modify their strategies and actualise their new knowledge can play an important role in learner growth (White & Frederiksen 1998). The three-stage model of Hattie & Timperley attempts to synthesise the difference between current student understanding, and where it should be, via a process driven mechanism of feedback. This concept integrates well with that of Abrams & Gerber (2013), who claim that feedback is the authoritative source of information for any change to be made between where the student's skills are, to where they should be. As feedback in Abrams (2013) model informs change only when the user's preferences are taken into account, it can be seen that defining feedback through a series of *why?* questions, can allow feedback to be presented to the user through various mechanisms, but only when the user is receptive to this information (Abrams & Gerber 2013).

While feedback forms a vital role in the accretion and synthesis of new knowledge for learners, a deeper level of feedback is required for a learner to engage more fully with educational content, as has been shown in this section. Being advised of how one has learned is not the full sum of the learning equation; comprehending, synthesising and reacting to new knowledge forms the bases of educational reflection; this is discussed further in Section 2.2.1.5.

2.2.1.4 In-Game Feedback

According to Abrams & Gerber (2013) video games can easily be described as a feedback loop. A player engages in a behaviour or activity in a game, the game responds with formative or summative feedback; the player observes the outcomes of their actions and according to their preference, goals, beliefs, enjoyment or flow, will accept, reject or modify a mental schema concerning their knowledge of what has transpired. This process is a continual and for experienced players, an instinctual part of how video games are played. In a sense, the feedback loop of video games is analogous to the iterative approach of formative feedback. Through continuous cycles of feedback, as long as the learner has bought into this feedback, reflection for the student can occur, if feedback is presented in an appropriate manner (Abrams & Gerber 2013; Abrams et al. 2018). Abrams & Gerber (2013) relate this form of feedback to Vygotsky's principals of proximal development. A player will engage in these behaviours to the extent of their abilities and current mental understanding of a situation. The feedback fed to the player

is the more authoritative agent, which allows the player to extend beyond their current zone of knowledge and this can take the form of the game itself, other players, tutorials etc. If a player shoots a rocket at their opponent and they score ten points, they have learned that points are awarded for shooting a player with a rocket. The learner is presented with the why of how they performed and were assessed (Hattie & Timperly 2007, where am I going?), along with the knowledge of how this relates to future iterations of their play updating a person's mental schema). This knowledge is provided via the game's feedback systems intrinsically. In contrast if the same player observes another player shooting an arrow which goes through an enemy's head, and thus gain 20 points, their proximal zone has been extended by new knowledge from a more competent or experienced agent, this time, another player. In both cases, the game has provided for the means of delivering the feedback, but for the former, the player learned internally from the game and externally from a player in the latter example (Abrams & Gerber 2013).

Murphy (2011) describes two varieties of feedback, both present in current educational and commercial video games. Video-games prove useful in describing feedback, as at the core of most video-games, lies a feedback loop. Short-term feedback, where the player is consistently aware of the outcome of their actions through visual, aural or haptic clues such that, secondary meaning of the feedback is not displayed. For example, a player may need to jump over holes in the terrain in a game. The visuals, audio and potentially physical feedback relay to the player whether they were successful. This feedback engages the player with the knowledge of whether their action was immediately successful. Holistic feedback expands upon this by attaching greater meaning to the action than just the binary success/failure. The game may relay to the player that they may have jumped over the obstacle, but instead they missed an opportunity to fall down the hole to pick up loot. The lesson learned for the player then is that while jumping some obstacles may be good, there may be opportunities for better rewards to fall through them. Holistic and short-term feedback may be both present during the moment the player performs an action, but the meanings behind them can alter the player's perception of the game (Murphy 2011). This altering of player's perceptions forms part of the reflective process detailed in Section 2.2.1.5.

2.2.1.5 Reflection

Reflection, as defined by Boud et. al cited in Edwards et al. (2013, p.3) is composed of "*Activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations.*" Reflection is therefore processes and activities which engage a learner with what has transpired, in a similar manner to feedback. Feedback and reflection can be viewed as complementary tools to one-another in that while feedback relies on a one-way process of absorbing new information from one's previous performance, reflection relies on the engagement of the learner on understanding what, why and how something transpired delivered via feedback (Mezirow 1990). While corrective feedback may aid a learner in understanding where their current knowledge is inadequate or incorrect, the value of a student reflecting on their incorrect mental schema may prove less than beneficial (Moreno & Mayer 2005). Reflection can therefore be viewed as another tool to aid the student's learning processes, when positioned carefully in an educational setting.

Reflection can be further described as transformative learning or as higher-order learning processes (Mezirow 1990). Both these terms importantly distinguish that reflection is a process which sits above the regular learning process. This distinction importantly underlines what Mezirow (1990) describes as reflexive vs reflective action. Reflexive action requires the learner to engage in a thoughtful and active process of analysing what course of action to take or what knowledge to take from a given situation given their current mental schemas. This opposes reflective action which relies on the learner to pause and engage in both forward and backward gazing introspection on the state of their own knowledge (Mezirow 1990). This distinction is important to bear in mind when considering learning models and how reflection and feedback are supported.

2.2.1.6 Reflection and Video Games

As mentioned previously, reflection is the process of people engaging with their experiences in an attempt to form new schemas about current and new knowledge. This reflection can occur as an activity within, alongside, or separate to the gameplay. While reflection is an important component of the Kolb (1984) Experiential Learning Model, a

relatively small subset of research has been undertaken in the space of in-game or post-game reflection and how it relates to the learning experience. Reflection can be stimulated via guided tutorials (Seale & Cann 2000) and smart tutors (Aleven & Koedinger 2002). Reflection has also proven to be effective learning tool in some cases when players can provide explanatory reasoning to game outcomes (Moreno & Mayer 2005). While these are positive results, the literature is still not at a consensus relating to the efficacy of reflection. Cowley et al. (2013) attempted to insert a social period of reflection into a serious game and found that students who engaged in social reflective practices such as group discussions and open-ended questions performed worse in learning retention tests. This corresponds with the results of Moreno & Mayer (2005) who designed a study which allowed players to explain their actions following a gaming session, which resulted in lower knowledge transfer for the students overall. The students who did not engage in any reflective practice, but instead were provided with explanatory feedback on their actions increased their knowledge retention.

Reflective experiences provided within a video game context can prove useful for learning but are not foolproof. While reflection is touted as an important process of higher-order learning, successfully engaging players in reflective and not reflexive practices is still open for exploration. The mechanisms through which reflection occurs are detailed in the following section.

2.2.1.7 Video Game Mechanisms Summary

The preceding sections have detailed important components of video games that may be applicable to the domain of literacy education enabled through technology. The mechanisms of flow are what drive players to continue in the experience, perhaps long after they would have disengaged with alternate activities. Feedback and reflection are mechanisms through which players are presented with new schemas of information, or modifications for old ones. Feedback as a cycle was explored, and the mechanisms through which feedback is disseminated either implicitly or explicitly was discussed. Reflection as the process of being presented with new information to change a person's schema on a concept was also discussed, with the tight interaction between feedback cycles and the context in which they are presented to the player. While the broad concepts presented in this section have detailed the mechanisms through which players

may enjoy, and make meaning of video games, the systems through which they are enabled are yet to be explored. The following sections will outline how these theoretical concepts are tied together, under the discourse of serious games.

2.2.2 *Serious Games*

Serious games are video games which attempt to provide an experience to a player, beyond that of pure hedonistic enjoyment (Charsky 2010). The discourse on serious games can be seen as a reaction to the 'edutainment' movement of the 1990s and early 2000s (Ma et al. 2011). Edutainment sees video games as a means to an educational end, crafting generic and often repetitive gaming experiences around educational content; edutainment has in some ways affected the reputation of video games for learning (Resnick 2004). Okan (2003) contend that edutainment attempted to sell video games as a pedagogical framework, and not simply as a medium through which the content could be delivered. It is therefore understandable that when several edutainment products failed to live up to their claims of effortless learning, the term edutainment gained a reputation as a lazy way to teach (Resnick 2004). Serious games can be seen as a reboot of the idea that games can teach, while also being engaging (Jarvin 2015). This outcome is achieved via a much closer alignment of educational games to pedagogical theory, empirical testing and academic and commercial research and development (Charsky 2010); the manner in which this is achieved is outlined in the following sections. Serious games, as described by Susi et al. (2007), are a mechanism to provide more immersive experiences for players, without sacrificing entertainment value. The factor of immersion is important to the understanding of the reason why serious games became a topic of discussion in the education sector.

	Serious Games	**Entertainment Games**
Task vs. rich experience	Problem solving in focus	Rich experiences preferred
Focus	Important elements of learning	To have fun

Simulations	Assumptions necessary for workable simulations	Simplified simulations processes
Communication	Should reflect natural (i.e., non-perfect) communication	Communication is often perfect

Table 3 Comparison of serious games to entertainment games

2.2.2.1 Forms of Serious Games

As described in Table 3, taken from Susi et al. (2007), serious games can be seen to be problem solving in nature with regard to tasks, compared to entertainment games which are designed to encourage experience over task. Again, the focus of the game too, are elements of learning over those of pure enjoyment. Notably, communication, described for serious games as being non-perfect, where it is up to the player to interpret instruction for the purpose of learning, over that of entertainment games, which seek clearly communicate almost all things required of the player (Susi et al. 2007). Serious games according to Ma et al. (2011) have a secondary aim beyond that of just entertainment; to teach, train, or simply to reinforce on or multiple skills (Susi et al. 2007). The desired outcome of serious games, are what Shaffer (2006) describes as epistemic frames, which are the collection of knowledge, skills and identities of a given topic, held by a group of people (Lieberman 2010). More simply, these frames are conceptual models of the real world, which can be translated from the game experience to the real-world. Furthermore, epistemic frames allow the immersion of a person via both a game or the real world about a specific piece of content, as they act as agents of the system by learning through active participation in both the game world, and the social construction around it (Lieberman 2010). It is the goal of serious games then to impart epistemic frames from a game context, and be translated into the real world (Lieberman 2010).

Serious games can be thought of as existing on a spectrum, according to Marsh (2011), defined by the dedication that the particular game pays to crafting these

epistemic frames into the core gameplay. More simply, serious games all aim to impart a secondary goal beyond that of play, (Susi et al. 2007), it is the manner in which the content of the game shapes the design and feel of the game that determines the form of the game (Marsh 2011). While serious games may be designed and developed with the intention of informing, and providing enjoyment, several studies have examined the efficacy of commercial of the shelf (COTS) games for the use in education (Charsky & Mims 2008). Notable examples include *Civilization III* (Charsky 2004) for teaching global history as well as *SimCity 2000* (Adams 1998) for teaching geography and urban development. While these studies have found significant improvement in educational outcomes, there exists a large caveat—that for COTS games to be worthwhile in educational contexts, the gameplay needs to match the curriculum well (Charsky & Mims 2008). Crafting a strong epistemic serious game experience requires a collaboration with both educators and game designers; the frames from the video game must translate to the real world. Potentially, designing an educational video game that players do not enjoy, and teachers do not find relevant is an ever-present threat to educational game design. While *Civilization III* may teach reasonable lessons about social policy, economic management and diplomacy, it would not be considered historically accurate, and so it may fail to fit perfectly as a history tool for educators. Educational systems are less likely to adapt to video games, than video games adapting to educational systems (Lieberman 2010). Therefore, a more reasonable approach would be the selective application of educational or commercial video games into specific courses. For example, using *Civilization III* to teach lessons on economic diversity, while ignoring content that falls beyond the scope of the required content. The creation of serious games exists on the spectrum of repurposing COTS games, to bespoke solutions that fit a specific domain. The following section details how serious games may be used for literacy education and refers to both custom and COTS forms of game.

Another factor for serious games is that of multimodal learning. Multimodal learning, as described by Jewitt et al. (2016) is the concept that meaning is made in a variety of ways. Serafini (2012) states that as the complexity of texts increases for the reader, so to should a shift from monomodal to multimodal education occur to help the learner more thoroughly understand these texts. Generally multimodal learning

describes learning as occurring through a variety of mechanisms and mediums, such that our understanding of things, and thus our learning of them, is filtered through multiple sources. Jewitt et al. (2001) describe some forms of multimodal learning in the classroom as occurring through mediums such as reading teacher body language (visual), intonation (aural), repetition through copying text (read/writing) and actions (kinaesthetic). While these modes of learning provide a rich multi-avenue approach to student learning, serious games can present their own form of these multimodal learnings, particularly those of alternate human to computer interfaces such as brain interfaces and haptic feedback (De Freitas & Liarokapis 2011). Even less revolutionary technology can be seen as aiding multimodality with regard to learning (Blumberg et al. 2013). Serious games present an opportunity for multimodal aspects of literacy education, through the fact that video games require several modes of interaction between the player and the game. Aspects of serious game design can be seen to present content in a multimodal manner, such as a deliberate feedback loop between a teacher or a student, along with the space and capacity for an activity to be iterated and practiced, including learner reflective periods (Anolli et al. 2010). Factors such as measurable and demonstratable progression of learning and mental models can also contribute to a multimodal manner of learning(Anolli et al. 2010). These factors will be explored in the following sections, and future reference to factors such as reflection, and feedback will be explored.

2.2.2.2 Serious Games for Literacy

Serious games have a long relationship with literacy education (Lieberman 2010; Gee 2005; Tjus et al. 1998; Sqire 2005). As previously demonstrated, literacy is a core component of many countries' adolescent education system; with video games being a core entertainment medium (Gee 2005). The intersection between the two mediums presents the opportunity to engage young people with a medium of entertainment that is present in their everyday life, with a core skill that will enable them throughout their lives.

Lieberman (2010) argues that when students become proficient in and enjoy a certain a game, they are more likely to read, or write about the game, re-engaging with the epistemic frame. The source text can occur within, or external to the game, the key

consideration being that students are more engaged to read, listen and write with a topic that interests them. According to Lieberman (2010), players who engage with *BioShock* on a regular occurrence, frequently encounter audio logs which guide them through the narrative of the story, and are found in a non-linear manner; it is up to the player to reconstruct meaning. Gee (2005, p.40) states that gamers are exposed to a *"multi-modal literacy par excellence"*. This is due to the highly dialogue driven nature of the game's narrative, with comprehension of the game world coming from audio logs scattered throughout the levels. Players can seek external sources to the video game in which to learn further about the game, particularly online. The proliferation of online and book strategy guides, forums and wikis, attests to the desire for game players to share their knowledge of the game and learn from others. This form of sharing can be seen as a form Vygotskyian social constructivism, where shared meanings are formed and distributed amongst groups of aligned learners (Chaiklin 2003). This however only applies to COTS games in the most part; bespoke games will not necessarily have this level of cultural impact to generate secondary texts in the wider gaming community. Using games for literacy requires educators to be active in the understanding of the video game and a knowledge of how to teach players the game. This presents the challenge whereby educators must become experts in newer technologies and modes of delivery, of which they may not have experience, or even interest in (Ray & Coulter 2010), this may be exacerbated by COTS not aligning to teacher obligations or interests.

A variety of serious games exist in educational settings which aim to increase the literacy of players and exist as COTS games with a highly specialised focus. Games such as *Reading Eggs* (Blake eLearning 2017) attempt to teach children in Primary School basic lessons of reading such as vocabulary, phonemes and object association with the alphabet through animation, sound, and basic gameplay. The purpose of *Reading Eggs* is to present literacy materials to a young player, through the medium of a video game, as presented in Figure 2-4. *Reading Eggs* presents the player with a variety of simple games, where literacy content is embedded into the game, and it is the purpose of the player to perform a reading activity to complete the game. *Reading Eggs* is a widely popular educational system, particularly in Australia, with over 16,000 primary schools subscribed to the platform in Australia alone.

Figure 2-4 *Reading Eggs* gameplay

Other games, such as *Grammatikus* (Sharp & Fitz-Gerald 2016) attempt to teach the player lessons in grammar as well as text analysis, punctuation and reading comprehension through a series of structured exercises, as shown in Figure 2-5. *Grammatikus* aims at an older audience of late-primary school aged children, and presents a large, more complex system for players to interact with. Players have a permanent character, with access to achievement tokens such as scoring, leaderboards, and character progression items. Players engage in a variety of activities where the player must solve grammar related problems via simple interactions with the game world. The challenges stem from the player's understanding and comprehension of the grammar presented, and gameplay revolves around choosing an action, such as swing the character's sword at an incorrect noun, and watching the result play out. Both solutions of *Reading Eggs* and *Grammatikus* offer a multimedia approach to teaching and testing basic literacy skills at different age levels.

Figure 2-5 ***Grammatikus*** **gameplay**

While both serious games are COTS, no formal research has been conducted to determine their efficacy in the school system. They remain educational products which are bought by schools on the presumption of their usefulness, with empirical evidence not forthcoming. The success of these games however does not appear to be reliant upon empirical testing; it would appear that schools adopt serious games before they have been rigorously tested (Cowley et al. 2013).

2.2.2.3 Barriers and Limitations

The current limitations of serious games also apply to literacy games, and in some cases, extend further, due to the peculiarities of the topic. A typical complaint of serious games is that they fail to teach what was required, or teach something different to the syllabus set by the educational institution (Susi et al. 2007). This presents a risk for educators, as failing to meet performance expectations set by educational institutions is to be avoided. While literacy games may aim for the same outcomes as the school curriculum, the way the game goes about scaffolding this may be in direct contradiction to teaching best practices for that school. Secondly, serious games may either fall into the 'too-fun' or 'too-serious' trap (Susi et al. 2007). If games provide too much enjoyment, and too little learning, the serious game has failed to meet its dual objectives of fun, and learning. So too for games which, may grimly adhere to the coursework, but fail to elicit any form of

motivation in the user at all; this too would be a failing of the serious game (Marfisi-Schottman et al. 2010). A further barrier impeding the progress of serious games for literacy is the overall cost, time, and expertise required to build successful serious games. As educators may not be game designers, commercial or research entities need to be involved in the process of designing, building, and monitoring the performance of these serious game systems (Susi et al. 2007; Hébert & Jenson 2017). Further work has been placed into the issue of serious games, and the desire for more empirically derived methods of game design (Steinkuehler & Squire 2014). Steinkuehler & Squire (2014) argues that there is little consensus in the literature concerning video games for play that codifies game mechanisms and their desired learning outcomes. This deficit in the body of knowledge shall be explored in later sections.

While serious games may promise to reignite student motivation and engagement, the body of evidence supporting net positive changes is not unambiguous (Wouters et al. 2013). While educational games have existed in isolation such as *Lemonade Stand* (1979) and *Oregon Trail* (1985), the thoughtful integration of these games into learning and training environments may have hampered educational gaming's success rates; instructor support toward serious games in general is an important component to the successful integration of serious games into classrooms (Ray & Coulter 2010). A recent example of such a disconnect was revealed in an Australian study, by the Department of Education and Communities. The Department integrated a serious game (called *MangaHigh*) into nine schools in New South Wales for a single term (Department of Education 2011). The program, according to Department of Education (2011), anecdotally increased the self-reported engagement and motivation of the students. As an example of structural problems relating to instructor support however, *MangaHigh* appears to fall to similar criticisms of other serious game systems. The NSW Department of Education (2011) recommended several alterations to the way any future serious game system should be rolled out to educational institutions.

- Firstly, the game was received well by students from both genders and across several high-school years; the game however did not have the full support of all educational providers, some believing that a lack of technical instruction for the teachers did not

allow the program to be fully utilised. It was in a sense of greater enjoyment and use to students, than it was for teachers using it as an administrative and diagnostic tool/

- Secondly, the game focused too specifically on one genre, with anecdotal evidence from students, particularly female, stating that a mix of game genres would be beneficial to the enjoyment of the game.
- Finally, the serious game content was perceived by the teachers as following the curriculum broadly, but not enough to allow full integration into teaching schedules. The educational providers found themselves integrating the game at the end of a module, as a validation tool rather than during the term as an educational tool (Department of Education 2011).

The case detailed is indicative of the findings of Wouters et al. (2013) and Meyer (2009) whereby serious game systems being adopted by educational institutions are not being used to their full potential: lack of training, as well as over- or under-generalization of the subject matter leading to potentially increased motivational outcomes for students, but poor adoption rates by teachers. The issue of a poor fit of content to the curriculum, being of most noteworthy concern, along with declining interest or motivation by the teacher to fully support the system (Becker 2007). The following section will detail two forms of serious games, and how they may relate to coverage of the curriculum in classrooms, and the way it is presented.

2.2.2.4 Complex and Mini-Games

Prensky (2005) states that a form of video games—called *complex games*—attempt to provide a complete educational content package, by integrating multiple game mechanics, stories, characters and settings. The aim of complex games is to emulate the commercial success of games designed for play rather than education. Since complex games are popular in society, especially with adolescents, complex educational games seek to emulate the mechanisms by which players enter the flow state or generally enjoy a game. Complex games have a history in traditional video game and serious game settings (Marsh 2011). It can be observed that the majority of games on the commercial market are comprised of detailed, and bespoke entertainment experiences (Prensky 2005). Complex games do find their way into the educational domain, with games such as *The Typing of the Dead* (a COTS game viewed in the lens of educational games

(Mozelius 2014)), with a focus on reading and touch typing (Modern Dream 2013). Further examples such as using *SimCity 4* to teach civics to high school students (Woessner 2015), prove that there is a place for complex games in the educational sector. It must be noted however, that for a complex game to be successfully integrated into the classroom, effort is placed on the teacher to adapt the curriculum and lesson plans to fit the static content of the game (Prensky 2005).

Mini-games as defined by Prensky (2008) are games which engage the player for short periods of time across a narrow set of gameplay mechanics. Prensky (2008) asserts that if educational mini-games are utilised as a means for teaching an educational curriculum, several benefits may be possible. Firstly, mini-games by their own nature are short and relatively straightforward to play. Cannon-Bowers (2010, p.2) states that mini-games "distil a complex learning concept into a small extremely targeted amount of game play". This effect means that engaging in a play session with mini-games during a class may be more realistic for students and teachers, as time and budget constraints are a concern (Prensky 2008; Prensky 2005). This is supported by De Jans et al. (2017) who designed an informant based design using mini-games for $17,000. Furthermore, mini-games have an easier distribution channel, being more easily accessible over the internet when compared to complex games, and are easier to string together, re-organise, and modify by the developer (Devisch et al. 2018). Prensky (2005) is somewhat critical of mini-games, describing them as "bubble-gum" and a potentially easy way out from designing games which attempt to engage the user in more complex fashions. Furthermore, Prensky (2005) argues that students wish to engage in non-trivial gaming experiences, especially in an educational setting. The criticism of mini-games does appear to be relegated to Prensky, and while he is a large influencer in the space of serious gaming, his opinions do not appear to have generated much discussion.

2.2.2.5 Mini vs Complex

It can be easy to view mini-games and complex games in a completely dichotomous relationship. Mini-games can be produced cheaply and target very specific learning topics, yet be subject to lazy game design principles and cookie-cutter production (Frazer et al. 2007b). Complex games on the other hand can produce rich, compelling narratives with meaningful game interactions, yet they can suffer from a lack

of funding or an over or under saturation of learning content (Prensky 2005). Another factor which is not alluded to by Prensky (2005) is that there exists a division between the cohorts of current educational games. Those that are vestiges from the first generation of educational games, derisively earning the epithet "edutainment", and subsequent generations of games which were a reaction to these games. The former category includes games like *Math Blaster, Carmen San Diego* and *Reader Rabbit*. These games can be perceived as shallow by academics, gamers and teachers alike and continue to act as the spectre of failure in educational gaming (Resnick 2004; Okan 2003; Charsky 2010). Bakker (2014) found the use of mini-games in a longitudinal study for increasing mathematical skills in students to be effective at encouraging students to engage better with their maths lessons, while also statistically performing better than their peers.

The argument against mini-games may appear compelling on first-glance. Prensky (2005) argues that mini-games are trivial, irrelevant to younger generations and lack the necessary depth to capture a player's attention or enter a flow state. While Prensky assumes that game players have moved on from smaller-games, the world has changed significantly since the mid-2000s. A new and dominant force in culture has emerged, which contradicts Prensky's assertions that children cannot engage with mini-games. The new trend is the rise of gaming mobile applications, or 'Apps'. While mobile games do not necessarily have to be considered mini-games, mobile games are typically played in shorter sessions than for PC or dedicated games console games. According to Euston (2014) in a sample of 60,000 Android smart devices, the average amount of time spent playing mobile games in 2014, was 51.8 minutes per day. Furthermore, 17% of the games time accrued per day were in the 'Brain and Puzzle' category. Extrapolating this data, the average American who plays games, spends 53.5 hours per year engaged in some form of puzzle, or brain-teaser game. Furthermore, Perez (2014) reports that time spent playing games on a handheld device had increased by 15% in 2014 when compared to time spent playing in 2013. Prensky's argument that children will not be entertained by games of a more trivial nature can now be called into question. While these games may appear to be *complex games*, the game mechanics are typically limited in scope, and interaction is typically limited to simpler input schemes. So too is the

average time spent playing these games, with an average session time of 5 minutes 35 seconds (Hwong 2016), representing very short interactions with these games per session. Although people may spend only several minutes per session on a game, they do continue to return to the game, and mobile games as a genre; an example of where games that can provide meaningful play experiences in a very short period. While educational apps may still be less popular than other non-educational apps, the significance of games with shorter session times has established itself in gaming culture. Recall the Hattie & Timperley (2007) Feedback Model presented in Section 2.2.1.3. The purpose of the model is for students to ask three basic questions:

- Where am I going?
- How am I going?
- Where to next?

These questions define how a learner, and indeed a player receives feedback from the system concerning their progress. Recall in Section 2.2.1.3 that feedback mechanisms form an integral component of learning (Lieberman 2010) and indeed play (Abrams & Gerber 2013). In the context of mini vs complex games, the application of these questions to the discourse is therefore important. For a player to know where they are going, they must contextualise where they currently are in the game. Mini-games have the advantage over complex games, in that the core gameplay experience is very clearly defined, with highly structured breaks in gameplay. For complex games, where the player is with respect to the game, and indeed the learning content, can be highly variable, and may be obscured through gameplay mechanics which are not enabling the player to consistently identify where they are in the content, and gameplay; more simply, complex games have a harder time contextualising an individual play session into the broader discourse of learning. Secondly, determining the state in which the player's knowledge of their goals, or 'where they are going, and where to next', is harder to express in complex games. With the competition of primary, secondary, and tertiary gameplay goals, players may not be always aware of what the goal of learning content is, and what is expected of them. While players may engage in a higher flow state with complex games, due to more in-depth goal setting, this may come at the expense of performing at the required level set out by their teacher.

It has been identified in the preceding sections that a lack of flexibility in complex serious games has contributed to several deficits in serious game coverage of in-class curricula. In comparison, mini-games may not suffer from an inflexibility to be re-arranged and repurposed for the specific activity; their deficit stems from the inherent simplicity of the gameplay, and the reductionism that may take place when designing them. With these points in mind, it is important to observe that little research into the efficacy of mini-games for literacy has occurred in the last ten years. In the following section, mini-games will be examined further in the context of compendiums, which attempt to alleviate the problems identified by Prensky (2005), and play to their strengths, as outlined in this section. The application of these compendiums for literacy education will be discussed.

2.2.2.6 Mini-Games and Structured Learning

Mini-games, as previously described in Section 2.2.2.4, are short and contained video games, with the intent of producing an interesting experience in a relatively reduced time-span. Mini-games have been previously used to teach a variety of topics, including history, numeracy, literacy and observational problem-solving skills (Annema et al. 2012; Korozi et al. 2012; Frazer et al. 2007b; Panagiotakopoulos 2011). The relationship between multiple mini-games in a structured context has however, received little scrutiny, beyond that of the work of De Jans et al. (2017), whilst designing a series of mini-games, did not evaluate their efficacy. While one mini-game may not be useful to teach an entire syllabus of information reliably, if that same syllabus was to be decomposed into discrete 'conceptual mini-games' as proposed by Illanas Vila et al. (2008), a larger portion of information may be digestible for a learner. Devisch et al. (2018) state that a conceptual model of mini-games is suitable over those of complex games because they can target more specific components of the learning process, over the final end goal of the desired behaviour. This leaves the problem of how to present these mini-games. There exist two possibilities for packaging educational content and then presenting the content as series of mini-games.

The first strategy is to examine the content for natural breakpoints in learning. This may not be too difficult. Consider a textbook, composed of chapters and again of sections of chapters. Potentially, a mini-game can be purpose-made to fit each sub-

section of a broader topic. For literacy, if students were to focus on vocabulary acquisition, each mini-game can be considered an atomic unit; one which does not require the input of any other unit. Topics such as reading comprehension, where a student must understand what they read, can be composed of layering of different inferential tasks. Each task is a separate component in making meaning of text, and yet contributes to the whole of the task (Perfetti & Adlof 2012). This is a common approach to reading comprehension education, and is well founded in pedagogical literature (Perfetti & Adlof 2012; Spörer et al. 2009; Yuill & Oakhill 1988; Alfassi 1998). Therefore, it is evident that a broad topic of can be decomposed to more discrete problems running across a theme, particularly in the literacy domain. Using the previous example, if the topic were split into three mini-games, each teaching the player how to make meaning of a passage of text, the player could participate in each video game in turn to learn the requisite skills inference and comprehension. At this point, the natural insertion of feedback mechanisms becomes apparent. Players are given an opportunity to reorient themselves with respect to goal setting, and also measure their expected performance against themselves and others (Hattie & Timperley 2007). It is important to consider that so far, the discussion of a compendium of mini-games considers games that are distinct from each other but are still thematically tied to the broader topic. These units are therefore considered 'conceptual mini-games' and can be viewed as individual—and atomic—units of learning.

To expand on the concept of atomisation, Illanas Vila et al. (2008) argue that for mini-games to be successful in their use as educational aids, the educational content contained inside the game must be conceptually concrete and consistent. Illanas Vila et al. (2008) propose a set of guidelines by which conceptual mini-games should follow.

1. The game mechanics contained within a conceptual mini-game should not change over the course of the game, to ensure consistency
2. The games should focus on a core, atomic component of an educational field, such that they can be treated as a 'learning object' Illanas Vila et al. (2008); and,
3. Conceptual mini-games should be composed of "infinite playability", where the difficulty of the game increases over time, ensuring a student never reaches a point where their ability outmatches the game's challenge.

The body of knowledge relating to the field of serious mini-games is small but growing. Studies such as Panagiotakopoulos (2011) attempt to use stand-alone mini-games to promote education in mathematics. While literacy and mathematics are somewhat disparate topics, the analogy of serious mini-games still holds true. Panagiotakopoulos (2011) utilised a conceptual mini-game, to target a specific target of educational need. Cannon-Bowers (2010) describe three cases of mini-games targeting the education and training of individuals, with generally positive results. Generally these mini-games can be considered to contain both a lower risk of failure for the learner, as well as the opportunity to engage in bursts of activity, with a primary focus on sub-goals rather than broad topics. This shorter play time while at the expense of longer flow periods, may enable greater long-term motivation, as players are not presented with confusing or protracted goals and game mechanics. In a structured review of 30 educational mini-games, Frazer et al. (2007b) found that educational mini-games in their current state lack of immersive gameplay and their over-reliance on prompting the user to perform actions. Of importance, was the conclusion made by Frazer et al. (2007b) that each mini-game they reviewed lacked any form of meaningful feedback for the player to digest, or the teacher to assess, and none of the games allowed periods of reflection during or after play. Finally, Frazer et al. (2007b) recommend that educational mini-games, either conceptual or not, would benefit from being placed in a broader context than as a stand-alone game. Frazer et al. (2007b) claim that none of the games, gathered the disparate learning resources into one comprehensive environment. The educational mini-games previously described are by their nature, atomic, which suggests they should naturally fit when arranged in series to meet course objectives. If the arrangement of educational mini-games into a "compendium" were to be undertaken, Frazer et al. (2007b) theorise that students may have a greater contextualisation of each mini-game as part of a whole.

As a serious game is intended for a purpose beyond that of pure enjoyment, the outcome of a serious game important. The defining question becomes 'what does this serious game want to teach?' The answer could be skills, cognitive knowledge or affective outcomes. For the different outcomes, different games, educational models and game characteristics can be focused on to ensure learning is achieved. For literacy

education, it has been identified that certain components of the curriculum are atomic in nature and rely on layering of strategies to gain vocabulary or practice comprehension. As has been discussed in this section, mini-games provide the opportunity for literacy education to remain highly focused on the goals and outcomes of the student, while also remaining flexible enough for the player to receive regular feedback on their performance and be enabled to contextualise their performance with respect to the literacy task.

2.3 Summary

The intersection of video games and education have in recent years produced an ever-increasing body of knowledge in relation to the science of educating. The units of this intersection are *educational games*, sometimes termed *serious games* or *edutainment.* Originally, the design of these games existed outside the realm of scientific methodology and were marketed by the individual company as 'potentially' educational. More recently, educational games have been the subject of intense scrutiny from academics and teaching professional in an attempt to resolve two key challenges in educational games and education overall. Firstly, the challenge of motivating people, especially children, to learn content that may not intrinsically stimulate them. The second challenge is to harness the attention of the learner on a game, or serious game and translate their mental energy into learning outcomes. Literacy is a key area in which many students may fall behind. The regional and macro costs of illiteracy are enormous, and a variety of remedial measures have been proposed to reduce the gap in educational outcomes. As discussed in Section 2.1.2, Australia, face serious challenges with respect to pre-high school literacy rates. These challenges are even greater in states such as Tasmania (due to high levels of illiteracy), and effort to encourage literacy remediation in Primary School aged students is deemed of importance. In a research capacity, broadening the body of knowledge regarding digital literacy education, particularly in Australia could form part of a move by educational bodies to encourage serious reform in this area. Targeting areas of the lowest literacy rate, such as Tasmania are part of an effort to close the gap between national cohorts of students. Along with this, successful studies across Australia, particularly for students most affected by illiteracy could

greatly aid efforts for educational reformers to conduct wider, and more generalisable studies, following on from this research's line of inquiry. Serious games have the potential to transform the teaching of literacy, and there exist several realms of serious game literacy experiences which have not yet been explored.

Currently, there exists a gap in the body of knowledge relating to mini-games in relation to the broader category of literacy serious games. While serious games and their form have been explored in depth, little research has paid attention to the size and scope of the games in question. Prensky (2005) argues that an important and oft-neglected component of educational video game theory is what he terms: "mini-games" and "complex" games. Specifically, mini-games are small, quick to play games which allow the player to engage with a narrow set of gameplays. Complex games on the other hand comprise everything else; longer, more involved games, potentially with deeper stories and secondary or tertiary goals. In both cases, adding significantly more content to the game or the text may not magically increase learning. What may increase learning is the role of feedback and reflective processes placed between portions of the content presented in a structured manner.

The role of feedback in educational games has been relatively under-explored. While authors such as Zebel et al. (2013) have demonstrated positive learning outcomes from the inclusion of between-game feedback, the role of elaborative reflection included with the corrective feedback has yet to be explored in detail. As mini-games can be run-through in matter of minutes, the role of feedback on player performance increases in importance, as the gameplay may not provide enough corrective or reflective space for the player to accrete new learning schemas. Similarly, a player may play a complex learning game for two hours and may only learn how to navigate the game world and 'win' the levels. For both cases, learning may not be provided the space it needs to be reflected upon. Abrams & Gerber (2013) provide the argument that continuous in-game feedback aids players in learning how to play a game. Similarly, semi-continuous feedback external to the game provide the space to critically reflect on what has transpired in the game.

The purpose of serious games is three-fold. Firstly, serious games should provide a meaningful and compelling reason for people to engage in play. Secondly, serious

games should harness the attention of player into learning some educational content. Finally, serious games of varying sizes, scopes and aims should provide players with an environment where they can consistently be engaged and reliably learn most, if not any, kind of learning material. Unfortunately, serious games are sometimes not meeting educational objectives, consistently disengaging players, and may not be capable of providing a reliable educational tool for learning content (Susi et al. 2007). The problem may lie with the competing aims of academia, educational institutions themselves, and most of all, the students. Even more likely, the way researchers frame an educational game as a catch-all solution for a broad topic. Furthermore, the development times, cost and developer experience required for educational games increase significantly as the scope of the educational game increases.

Small in-roads have been made into the topic of mini-games, particularly (Korozi et al. 2012; Illanas Vila et al. 2008; Frazer et al. 2007a). Described in Section 2.2.2.4 educational mini-games are rarely framed in a positive manner as a solution to the serious game problem (Prensky 2005), and in fact are rarely discussed in the literature at all. This fault does not lie in the manifest unsuitability of mini-games as an educational tool, merely the desire for researchers to produce large and complex systems to cover more educational ground. If the smallest unit of educational gaming is the mini-game, would the resultant unit of learning be 'mini' as well? Variety in gaming experiences drives people to play new games over the course of their life. But what if mini-games were to exist not as a stand-alone 'trivial' piece of entertainment, but as a structured series of learning tools, punctuated by assessment tasks? A topic, rather than being the focus of an entire educational game, could therefore be reduced to discrete mini-games, structured in a series which attempted to provide the same breadth of content as a complex game. Each game's educational content would flow into the next, with structured spacing of feedback and reflective measures placed between them. Furthermore, these games could be re-arranged by educational professional to suit the bespoke requirements of the course, swapping out and re-assembling the games to suit a new requirement. These characteristics constitute a novel and unexplored avenue of research, with respect to literacy serious games.

The following section presents several research questions, which are the natural conclusion of this literature survey, and focus on the use of mini-games for the use of reading comprehension education.

2.3.1 *Research Aim and Questions*

The general aim of this study is to explore the relationships between mini-games, feedback, and assessment with respect to literacy, and whether this relationship can generate positive learning outcomes for learners. More formally, the over-arching Research Question can be summarised as:

Research Question (RQ): *To what extent can reading comprehension skills in students be strengthened by employing formative feedback within a system of structured mini-game experiences, and can this system satisfy the needs of students and teachers.*

To answer this research question, many sub-parts can be deduced which formulate components of the ***RQ.***

Sub-Research Question 1 (SRQ1): In what ways do playing a system of mini-games increase reading comprehension performance when compared to traditional presentation of these activities?

It is expected that providing a system of serious games with the intention of generating a distinct and positive literacy outcome will perform better than the traditional form of teaching this content. While literacy is a key area of development for children and adolescents, several factors on teaching literacy via serious games has not yet been quantified. Firstly, serious games have been mostly analysed homogenously, with the assumption that the specifics of the educational game are less important than the actual educational content. While a balance needs to be struck between educational content and gameplay, due care to the specific mechanics of the game must be exercised Kiili (2007).

Secondly, the role of playing serious games iteratively has not been explored in detail. The work of Zebel et al. (2013) found that the inclusion of a period of reflective feedback and assessment between game sessions can improve player's game scores, but not their test scores. Due to the nature of their study which utilised a complex game, Zebel et al. (2013) argue that the role of in-game feedback processes may affect how

intra-game testing and reflection performs. This assertion is supported by Abrams & Gerber (2013) who argue that games provide continual feedback loops which may negatively affect inter-game feedback, thus decreasing learning outcomes. More simply, players may get better at the game, but may not transfer that skill to external contexts.

Sub-Research Question 2 (SRQ2): In what ways do playing a system of mini-games, with variable periods of inter-game feedback, increase reading comprehension skills?

It is expected that a system which provides periods of assessment between structured serious game experiences can increase the reading comprehension of learners. This is due to the nature of assessment as a form instructional support, which attempts to scaffold the learning activities with systems that aid learners. Formative assessment is one such system of instructional support. Stiggins (2002) states that regular periods of formative assessment provide opportunities for students to observe their own state of knowledge, reducing the time between learning and formal assessment of knowledge using summative assessment techniques. Furthermore, Black & Wiliam (1998) states that providing students with formative assessment can reduce the desire for an increase in rewards based on 'correct' answers, and engender a desire to 'comprehend' rather than be 'correct'.

The role that formative assessment plays in serious games has been an under-evaluated topic of educational game research, and yet the role of elaborative feedback in learning outcomes is well known. When providing learners with feedback that elaborates on correct or incorrect answers, educational providers or systems can benefit from increases in learning. The link between elaborative feedback and video games is however poorly documented. Furthermore, the link between regular and iterative formative assessment within the scope of serious games is also unexplored. It is known that elaborative feedback can induce greater learning outcomes, (Moreno & Mayer 2005) and (Zebel et al. 2013), but further research needs to be undertaken to determine the correlations between these factors. Understanding the role in which formative feedback plays on the literacy outcomes of serious games will be critical in determining futures modes of serious game assessments.

Sub-Research Question 3 (SRQ3): What is the relationship between motivation, game experience, and reading comprehension skill acquisition, when reading comprehension activities are delivered via mini-games?

By utilising techniques of elaborative formative feedback and combining them with regular game-experiences, insights into the nature of the inter-game and intra-game feedback cycle for serious games will be explored. It is hypothesised that providing students with reflective spaces to re-engage with their learning, following a serious game experience, will increase their overall enjoyment of the activity and motivation to perform similar activities. The assertion is based on the work of Mezirow (1990) who states that reflection allows students to engage back with educational content actively. The role with which game achievement translates to external achievement has not yet been fully explored. Zebel et al. (2013) found that game scores were not positively related to test scores when comparing complex serious game assessments. This may be due to the fact that complex games have a longer in-game feedback cycle, which may make learning outcomes harder to reflect upon externally. It is hypothesised that the mini-games may provide a stronger correlation between game-scores and test-scores, since the period between in-game feedback and external feedback is low.

Sub-Research Question 4 (SRQ4): How can a system be designed which allows the structuring of educational mini-games, formative assessment and elaborative feedback which satisfies the needs of teachers, students and game designers?

Designing a system to create a structured learning experience which captures the necessary pedagogy and educational content is no small feat. Currently, there exist a wide variety of educational game experiences which encapsulate a broad range of serious games, educational content, and structured feedback. Combining these systems together in a consistent approach, with the particular focus on mini-games and literacy is yet, unattempted. Furthermore, a wider focus on the feasibility of generating and implementing educational systems as a whole is lacking, particularly amongst serious games.

Answering these questions will be an important milestone in the serious game discourse, paving the way for better, more fun and original serious game experiences.

3 Methodology and Tools

This chapter will present details relating to the methodological process used through the course of this study. Details relating to the research question will be presented, along with the process relating to requirements elicitation, testing procedures, and the way the efficacy of the system would be judged. This chapter will present the tools, and justification for their use in this study, and is considered the first Chapter of the methodology, which is broken into three Chapters. Firstly, this chapter will explore the information needs of this study in Section 3.2, followed by the philosophical underpinnings of this research in Section 3.3. Section 3.4.1 describes the chosen behavioural methodology and begins exploring its application to the research question. Section 3.4.2 describes the experimental design of the study, which informs the following chapter. Chapter 4 will present the application of the tools and preliminary modelling presented in Section 3.4.1, specifically behavioural modelling (Section 4.2) and the subsequent game design resulting from this modelling (Section 4.3), in the context of designing a serious game for literacy. Chapter 5 will detail the experimental phases and details used in this study, concluding with a presentation of the way data was empirically collected. Specifically, Section 5.1 is concerned with the design of the intervention, Section 5.2 concerns itself with participant selection and Section 5.4 defines the manner in which data is collected and how.

3.1 Research Questions

The general aim of this study is to explore the effect of mini-games with structured feedback on reading comprehension levels for students due to its significance in learner outcomes, as discussed in the preceding Chapter. The Research Question (RQ) can be defined as:

RQ: *To what extent can reading comprehension skills in students be strengthened by employing formative feedback within a system of structured mini-game experiences, and can this system satisfy the needs of students and teachers.*

The following sections will outline the information needs required to effectively answer this question.

3.2 *Information Needs*

Answering the research question, and its constituent sub-parts is the primary focus of this study. Therefore, it is imperative to determine the informational needs, tools and methods to correctly address these questions. The following sections will outline the informational needs required for the individual sub-questions, and the manner in which it collected.

3.2.1 *Sub-Research Question 1*

Sub-Research Question 1: In what ways do playing a system of mini-games increase reading comprehension performance when compared to traditional presentation of these activities?

To answer this sub-question, the current state of an individual's reading comprehension performance needs to be measured at least twice. Reading comprehension can be measured effectively via multiple-choice quizzes, and this form is most commonly used in the educational (ACARA 2017b), and academic domains (Rupp et al. 2006; Keenan et al. 2008; Sarroub et al. 1998). To quickly and accurately measure the state of change with respect to reading comprehension skills, an electronic test is deemed most appropriate (Mayer et al. 2014).

The second component of this sub-question is the reading comprehension activities themselves. This study is focused on the question of whether reading comprehension activities can be gamified, and how the presentation of these games can affect reading comprehension skills. To measure the efficacy of gamified reading comprehension activities, a baseline is required to determine whether students exposed to gamified reading comprehension have a significant improvement in skills when compared to students exposed to traditional reading comprehension activities. The comparison between groups in studies of serious games is well established, and is a dominant empirical design in this space, particularly quasi-experimental and randomised control trials (Koops & Hoevenaar 2012; Connolly 2012a).

3.2.2 *Sub-Research Question 2*

Sub-Research Question 2: In what ways do playing a system of mini-games, with variable periods of inter-game feedback, increase reading comprehension skills?

The purpose of answering this sub-question is to determine whether a system of mini-games that present formative feedback at regular intervals between gameplay sessions significantly improves reading comprehension skills when compared to mini-games without this feature. To measure the changes in reading comprehension skills, a pre/post-test will be employed, using the tools described in Section 3.4.2.1. The game system, which forms the independent variable of the study is described in greater detail in Section 4.5.

3.2.3 *Sub-Research Question 3*

Sub-Research Question 3: What is the relationship between motivation, game experience, and reading comprehension skill acquisition, when reading comprehension activities are delivered via mini-games?

The purpose of this sub-question is to examine the link between motivation, game experience, and reading comprehension skill acquisition through the course of this study. To measure these factors, several tools are required to collect subjective, quantitative factors (game experience and motivation) and objective quantitative factors (reading comprehension skills). A well-established measure of gathering the subjective experience of video game players after a play session is by utilising the Game Experience Questionnaire (IJsselsteijn et al. 2013). This questionnaire is a calibrated measure of player's experiences during a gameplay session, and provides a method to transform these experiences into a metric of engagement. A post-study survey will be used to measure the motivation and satisfaction of players following exposure to the game; this survey will obtain information pertaining to how motivated students felt with respect to the serious game, along with demographic information. This survey is expanded in detail in Appendix III.

3.2.4 *Sub-Research Question 4*

Sub-Research Question 4: How can educational mini-games be delivered such that they satisfy the needs of students, teachers, and game designers?

This question primarily concerns the usability of a system designed to answer the three preceding sub-questions. To answer this sub-question the opinions and expertise of the two identified stakeholders is essential, detailed further in Section 4.1.2. As the participants of this study are primarily students, with teachers forming a secondary participant cohort, the needs of these two groups are a primary consideration. As described by Pagulayan et al. (2002) and de Freitas & Jarvis (2006), requirements elicitation from users, using techniques such as workshops, and semi-structured interviews can generate highly relevant information concerning the design of a game. For the student cohort, a serious game needs to meet the criteria of enjoyability, replayability, and ease-of-use (Gee 2005). Enjoyability and replayability fall somewhat under the domain of personal preference, but certain game mechanics have been identified as boosting the success of these two criteria (Breuer & Bente 2010; Annema et al. 2012). Allowing students to voice their opinion on how the game should look, play, and feel has been well established in serious game design, and techniques such as in-class workshops have been successfully employed (Annema et al. 2012; Poels & Annema 2012). For the teacher participants, the usability of the system is a primary concern, as described in Section 2.1.3, for administering any educational game, helping students with their progress, and assessing the students' performance after the completion of the educational game; teachers require a well-designed and intuitive system. As Becker (2007) claims, the need for teacher input in the design of any serious game system that will be used in their classrooms is paramount. Semi-structured interviews allow teachers to voice their opinion on how a serious game system can be employed in their classroom, while also prompting them to provide specific information on required and desired features of the system. The outcome of this interview process is for researchers to be fully aware of the requirements of the system from a teacher perspective, using qualitative analytical techniques such as thematic analysis (Aronson 1995); such

techniques have been employed in the serious game space successfully (Bakar & Nosratirad 2013; Hess & Gunter 2013; Forum et al. 2005).

3.2.5 *Information Needs Summary*

To summarise the informational needs of each sub-question, and the tools and methods chosen to collect said information, a table of information needs has been generated (Table 4)

Questions	Information Needs	Data Collection Source	Data Tool
1	Reading comprehension skills	ACARA Improve service	Pre/Post-test
2	Reading comprehension skills	ACARA Improve service	Pre/Post-test
3	Student experience	Likert scale of experiential factors	Game Experience Questionnaire IJsselsteijn et al. (2013)
	Student motivation	Likert scale of motivation factors	Post-Test Survey
4	Teacher Requirements	In-person recording and notes	Semi-Structured Interview
	Student Requirements	Teacher Summaries	Teacher delivered in-class workshop
	Student Survey	Demographic Survey	Post-Test Survey

Table 4 Summary of Information Needs

The following sections outline how the identified information needs inform the selection of a relevant methodological framework, drawing from the experiences of similar empirical studies.

3.3 Methodological Philosophy and Framework

Due to the empirical and experiential nature of existing research into reading comprehension, and educational video games respectively, the need to capture both qualitative and quantitative data is required. Due to the qualitative and quantitative nature of the data required to answer the research question, a research design that is flexible in the management of these two forms of data is apparent. Mixed-Method designs are consistently employed in this research domain (Mayer 2012). As a framework, Mixed-Methods proposes the pragmatic need to mix data from multiple sources, and data collection techniques to answer a research question (Creswell & Clark 2007). Mixed-methods act as a middle way between strict quantitative and qualitative research frameworks, and espouses the need for a pragmatic approach to human-centred research (Johnson & Onwuegbuzie 2004; Creswell 2013; Johnson & Onwuegbuzie 2004).

Research into serious games has had a long, varied, and fragmented history of research designs according to Westera et al. (2008). Westera et al. (2008) claims that an overriding factor in serious game research designs is the lack of consensus on how best to design, employ, and measure serious games. More recently, the need for tools and methods which encapsulate participant experiences has become an important factor in the design of research studies in the serious game space, according to Mayer (2012). The need to synthesise user requirements into the design of serious games, and subsequently evaluate them from a qualitative perspective is becoming of greater importance in this space (Mayer et al. 2014). This study therefore attempts to align with the new trend of serious game research, which focuses on standardising serious game research methodologies, to improve the problem of validity and replicability (Girard et al. 2013), and can be said to follow the precedent demonstrated by Mayer et al. (2014).

As previously described, the needs of the user form an integral component of this study; a mixed-method design thoroughly encapsulates the needs of this study, and is well supported by previous research in this space (Mayer et al. 2014). Philosophically, this study requires a flexible framework for designing a system that caters to the needs of different cohorts, while remaining true to the guiding educational principle of constructivism: (Duffy & Jonassen 2013). Pragmatism is ideal for adapting to the specific needs of a research problem, particular to education (Feilzer 2010). As constructivism aims to place the learner at the centre of the learning process (Duffy & Jonassen 2013), pragmatism places the needs of the research question at the centre of the research (Biesta & Burbules 2003). Pragmatism draws on the strengths of interpretivist design to value the subjective opinion of the learner, while quantitatively deducing statistical significance from experimental interventions (Biesta & Burbules 2003). This approach conforms to the pragmatist design philosophy chosen for this research. The serious games space consists of a wealth of research designs, including quasi-experimental, user-centred, and qualitative explorations (Girard et al. 2013). Many of these research designs do not have a core focus on teacher requirement elicitation, which as identified in Chapter 3, is a significant barrier to the positive outcomes of a serious game intervention (Baek 2008). The Mixed Methods methodology will be composed of a Sequential Exploratory Design (SED) (Creswell et al. 2003). SED is suitable for educational research (Ivankova et al. 2006); the design is capable of capturing both quantitative and qualitative data (Ivankova et al. 2006). More specifically, the design incorporates a period of qualitative data collection, which informs subsequent quantitative data collection (Doyle et al. 2009). SED allows researchers to capture quantitative data on the activity, and supplement the knowledge gained with qualitative analysis of subjective opinions (Ibáñez et al. 2014). The former being used to determine whether quantifiable changes within the participants were present, and the latter, to further elaborate, discuss, and support the findings by the using participants' subjective opinions. To ensure the validity of the system, and the study, a structured series of design and experiment phases are required. These phases are described in the following sections.

3.4 *Methodology and Tools*

The process for designing a serious game for literacy, presented in this study, follows a two-phase process; depicted in Figure 3-1. Chapter 4, and Chapter 5 will detail the design, and experimental process used for this study. This chapter will only describe the tools used throughout this study and provide justification for their use and appropriateness. Chapter 4, and Chapter 5 will present Phase 1, and Phase 2 respectively.

Figure 3-1 Study Design

3.4.1 *Phase 1: System Design*

The following sections will describe the tools that were employed throughout this stage. For purposes of providing an overview, Figure 3-2 provides a representation of these tools and their progressive development in this study.

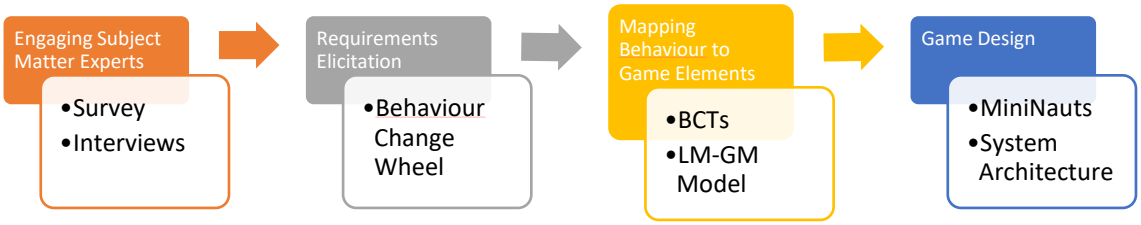


Figure 3-2 System Design Process

3.4.1.1 Engaging Subject Matter Experts

The System Design phase gathers, outlines, and implements the requirements of relevant stakeholders, for use in this study. Requirements, in the context of system design, are defined as necessary components of the system that satisfies explicit criteria imposed by stakeholders (Goguen & Linde 1993). These requirements in the form of expected constraints and required components were elicited through a series of stakeholder consultations; details of these consultations are described in the following.

Stakeholders involved in educational theory and practice, along with game design experts were engaged with throughout this study.

3.4.1.2 Design Methods and Standards

3.4.1.2.1 Human Research Standards

The primary aim of Phase 1 was to design a system in which reading comprehension could be positively enabled, through serious games. As the identified cohort for this study involved minors under Australian law, it was necessary to consult national guidelines pertaining to human research standards in Australia. These standards derived from the Australian National Health & Medical Research Council (NHMRC) (Health & Council 2017), and are detailed below.

- **Research must fully encapsulate the research question**
    - The focus of this study relates to the individual reading comprehension skills of students, and whether they can be improved using a system of mini-games. Therefore, the proposed system must capture student reading comprehension levels, provide mini-game experiences, provide feedback experiences, and finally measure whether a significant change in reading comprehension levels occurred. The NHMRC mandate that superfluous data collection, and methods that go beyond the scope of the research are not permitted (Health & Council 2017).
- **Is of Low Risk to Participants**
    - As with all human centred research, the risk to participants should be a prominent factor in designing any system which interacts with humans in an experimental design (Halai 2006). Due to the nature of this research, interactions with minors is also considered a necessity. The system therefore needs to be designed such that it satisfies the needs of the research question, while also presenting the smallest risk profile to participants.
- **Is Beneficent to All Participants**
    - As a counterpoint to risk analysis for any human centred research, the expected beneficence to participants should be a careful consideration in

the design of any system. The NHMRC present beneficence as the need to: "provide for the child or young person's safety, emotional and psychological security, and wellbeing" (Health & Council 2017).

- **Is Feasible to Use as an Intervention Tool**
    - The system, as defined by the research question needs to capture reading comprehension levels, provide structured gaming periods, and satisfy the needs of students and teachers. The system therefore has the dual needs of being a robust tool for administering and measuring an intervention to individual students, while also satisfying the needs of student learning and enjoyment, and teacher needs of standardised testing and student performance monitoring.

3.4.1.2.2 Software Development Lifecycle

The system described in the following section was iteratively developed using agile software development principles; a process whereby requirements, and the subsequent implementation of the software evolve over the life-cycle of the project (Cockburn 2002). This process afforded all stakeholders the opportunity to be aware of the current state of the system and have an opportunity to provide valuable critical feedback on the system. This process follows the precedent of Annema et al. (2012), whereby participants are given some guiding influence over what they would like to see in serious game experiences.

3.4.1.3 Requirements Elicitation

Phase 1 was initiated with a review of domain specific knowledge through consultation with stakeholders. The purpose of this review began with a dialogue with educational providers interested in committing to the study. A survey of teacher needs was generated; the questions of this survey are presented in full in Appendix IV. The content of this survey attempted to cover a breadth of topics; and aimed to elicit the professional opinions of the teachers on a variety of topics. These topics included barriers to literacy (Section 2.1.3), where the greatest deficit in literacy skill for their students lay (Section 2.1), and the practical manner in which they administered learning content to their classes. The purpose of the last point is for the designed system to closely

emulate how the teachers would administer content in their class to reduce disruption, and enable better teacher adoption; the manner in which teacher adoption of technological systems in classrooms affects the outcome of the intervention is described in Section 2.1.4. The survey further explored teacher opinions on video games, how effective they were, and what they believed their classes would appreciate in a serious game system. As previously described, the needs of the user are paramount in designing serious game systems (Nicholson 2012; Wood 1997; Pagulayan et al. 2002). Due to the restrictive nature of adolescent research in Tasmania, access to student participants was more easily mediated through teachers, rather than directly with the researcher, particularly in the early stages of this research.

The consultations previously described, aligns with the SED outlined previously and positive outcomes of stakeholder consultations has been identified by surveys of the extant literature (Backlund & Hendrix 2013). The purpose of eliciting requirements from teachers as the first phase of the study design is two-fold. Firstly, integrating users into the design of any serious game is considered best practice, and is well-supported in the literature (Pagulayan et al. 2002; Poels & Annema 2012). Secondly, as this study aims to explore the space from a teacher perspective as part of SED, the needs and requirements of teachers as participants of this study and users of the resultant system is of importance. It has been identified by Baek (2008) that teachers suffer from several barriers to integrating serious games into the classroom. These barrier range from applicability of the content, setup time being too lengthy, and the lack of direct measurable outcomes of students after interacting with the system (Baek 2008).

Whilst stakeholder consultations provide meaningful constrains, requirements, and suggestions for a serious game for literacy education, the capacity to translate those requirements, constraints, and suggestions into a meaningful video game requires careful consideration. To fully translate these requirements, a methodological framework was required which allowed the researchers to analyse these requirements, and the expectation for the final serious game's impact on student outcomes and behaviour.

3.4.1.4 The Behaviour Change Wheel Methodology

The Behaviour Change Wheel (Michie et al. 2014) is an aggregate of 19 behavioural frameworks that have been identified as supporting behaviour change; these frameworks alone however do not provide comprehensive coverage of extant literature according to (Michie et al. 2014). The BCW provides intervention designers with a systematic process for designing and planning a consistent strategy for behaviour change in a target audience. The BCW was expressly designed so that it can be applied to a diverse range of groups or individuals, based on the COM-B model of behaviour. The COM-B model aims to identify the components that lead to a target behaviour. To determine the applicability of each intervention function, a process of assessment was also developed by Michie et al. (2011). This assessment component is called the APEASE criteria. The following sections will describe the APEASE criteria and COM-B model in greater detail

3.4.1.4.1 The Behaviour Change Wheel

The Behaviour Change Wheel can be represented as a circular model with each step progressing further from the hub of the circle to the outer rim. The model is visualised in Figure 3-3.

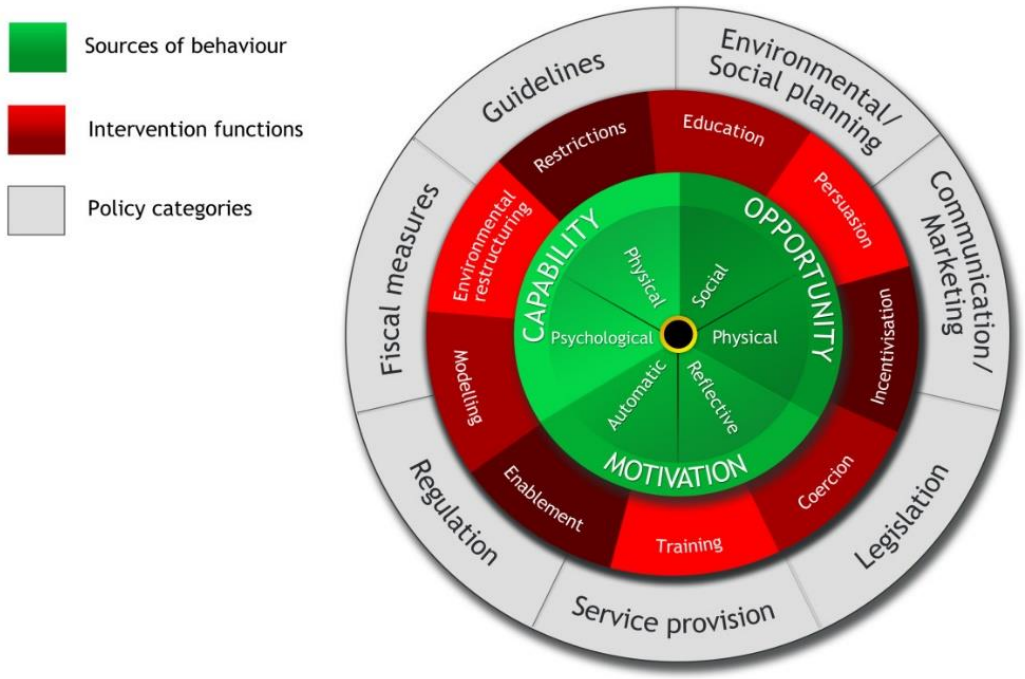


Figure 3-3 The Behaviour Change Wheel

The centre of the wheel is composed of the COM-B model, which represents the source of the behaviour identified, this is described further in Section 3.4.1.4.2. The second layer represents the way in which the targeted behaviour can be supported or reframed by intervention functions. The final layer represents the way in which these functions can be framed in the wider community or environment via policy.

3.4.1.4.2 COM-B Conditions

The COM-B model (Michie et al. 2014) aims to represent the conditions by which behaviour is generated in individuals. As previously described, the BCW relies on the COM-B Model at the centre of the wheel, such that a fundamental understanding of the wheel can be arrived upon before any intervention function is developed (layer 2) or supported (layer 3). The following table described in Table 5 describe the COM-B constituent components.

Component	Condition	Description
Capability	An individual's ability to engage in the activity	
	Psychological	The current cognitive ability of the individual to comprehend the activity
	Physical	The physical ability of the individual to engage with the activity, via their strength or dextrous skill
Opportunity	An individual's potential to engage with the activity	
	Social	Whether an individual is supported by their environment through cultural or social norms and expectations
	Physical	The individual's environment
Motivation	An individual's desire to engage with the activity	
	Automatic	The impulsive or unreasoned impulse towards the behaviour
	Reflective	The individual's mental model plans or evaluations of a situation

Table 5 COM-B Model

As can be observed the COM-B model is composed of the Capability, Opportunity and Motivation components that enable Behaviour. The figure presented in Figure 3-4 details the process by which the three Components affect an individual's capacity to engage in a behaviour. As such, any and a combination of each Component can have an impact on the behaviour of an individual.

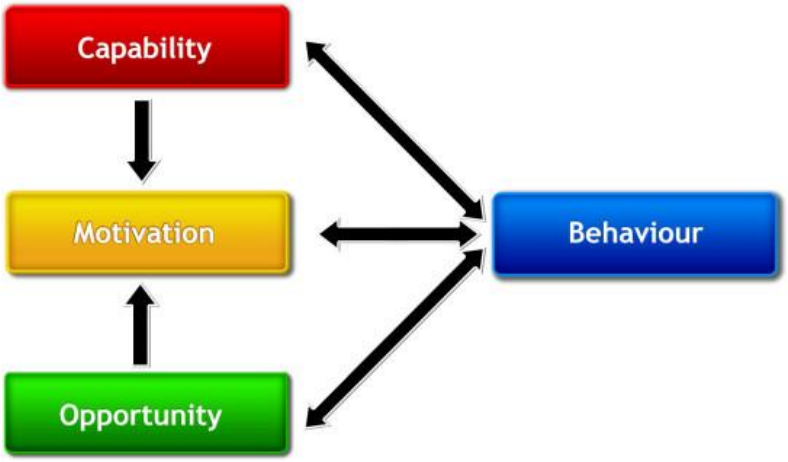


Figure 3-4 COM-B Model—Taken from Michie et al. (2014)

This model is used extensively throughout the design phase of this study. Through the BCW, a model of where problem behaviours lie in participants, and how these behaviours can be addressed with target behaviours is enabled.

3.4.1.4.3 Process of Designing Interventions

The BCW provides a coherent and well-supported model for mapping various factors of the environment and an individual's propensity to engaging with a certain behaviour (Michie et al. 2014). The BCW however is intended as a procedural document, which aims to facilitate the design of behavioural interventions (Michie et al. 2014). The three layers of the BCW provide a three-step process for identifying factors of behaviour, how the behaviour may be supported and finally how the behaviour can be framed via

policy (Michie et al. 2014). Figure 3-5 describes the process by which the BCW is applied to a theoretical behaviour.

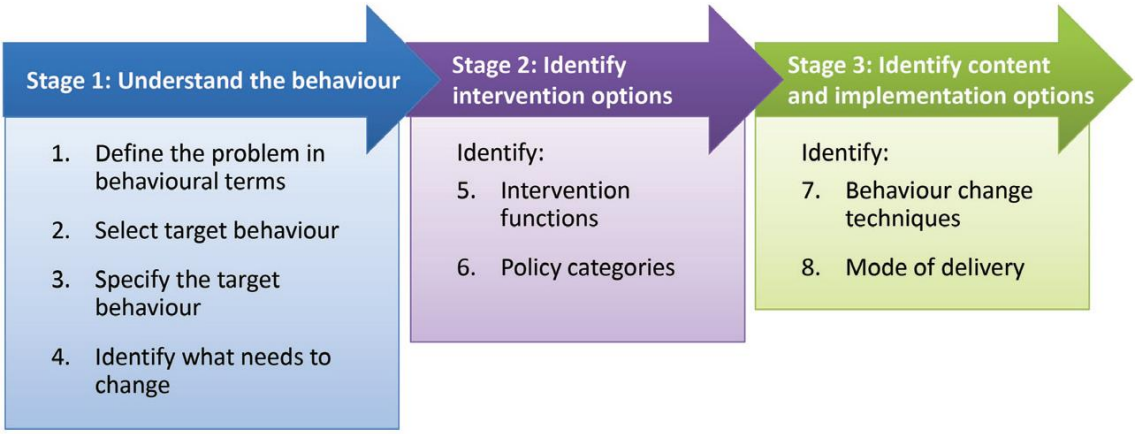


Figure 3-5 BCW Process—Taken from Michie et. al. (2014)

The BCW process aims to ensure that intervention designers properly define the target behaviour, such that the behaviour being intervened against is reasonably assumed to be the leading suspect for an individual's activity (Stage 1). Following this, intervention functions are identified by the designer and evaluated using the APEASE criteria to determine their applicability in the given scenario (Stage 2). Finally, the APEASE criteria are used to frame the intervention function in terms of its mode of delivery to the individual. Each mode of delivery is systematically evaluated for its APEASE applicability (Stage 3).

3.4.1.4.4 APEASE Criteria

The APEASE Criteria are a method for evaluating the *effectiveness* and *practicality* of the intervention function in actionable terms (Michie et al. 2014). The APEASE Criteria are described in the following table (Table 6), and will be used extensively throughout Chapter 3 to evaluate the state of each stage of the BCW of this study. Criteria used in this section refer to the work of Michie et al. (2014), and have been used in accordance to the methodology provided in that work.

Criteria	Description
Affordability	The intervention function in financial terms and how likely an individual/group can afford the function.

Practicability	The ability of the designer to implement the function in the environment with a reasonable expectation as to effort.
Effectiveness and cost-effectiveness	How impactful the intervention will be when measured in a real environment, along with the ratio at which expenditure and value are related.
Acceptability	How the targeted intervention is viewed by the community, experts, and society in terms of acceptance.
Side-effects/safety	How likely the intervention is to cause un-intended consequences for the target audience, along with the overall psychological, social and physical safety of the target audience when exposed to the intervention.
Equity	The likelihood that the intervention is to provide disparity in a positive or negative direction of an individual/group's standard of living, health, social status etc.

Table 6 APEASE Criteria

The APEASE Criteria form an integral part of the evaluation of functional changes to a person's behaviour, in a consistent and holistic framework. The application of the APEASE Criteria therefore form a core process of the BCW framework, and the application of this process is detailed in Chapter 4.

3.4.1.4.5 Behaviour Change Wheel Summary

The Behaviour Change Wheel aims to provide a consistent and process driven methodology for designing interventions. In Chapter 5, details of how the BCW was applied in the context of literacy education mediated through serious games will be presented.

3.4.1.5 Behaviour Mapping to Game Elements

Comprehensive behavioural mapping is used to determine the best way behaviour could be modified via the BCW, a crucial component of this process is translating behavioural mapping of target behaviours, being supported by game mechanics. Whilst the BCW process allows researchers and intervention designers to describe problem and

target behaviours in detail, the methodology does not present the way these behaviours can be modified in detail. What is required, is the ability to map learning methods and objectives to game mechanics. The BCW process described in Chapter 5, produces a matrix of possible behaviour change techniques, and how they would relate to the target. The manner in which these techniques are translated into game mechanics is an area of active research and exploration (Baranowski et al. 2008; Girard et al. 2013). More specifically relating to this research, Arnab et al. (2015) argues that there is a lack of common vocabulary for game designers and educators, when discussing serious games and their mechanics. Westera et al. (2008) argues that whilst many serious game design methodologies are adequate for use in evaluating whether a serious game and its constituent game mechanics are appropriate, they do not address the relationship between learning constructs and gaming elements. The way game mechanics are tied to educational constructs is still in flux; the following sections outline possible approaches to mapping serious game elements to the outputs of the BCW.

3.4.1.5.1 Gamification

Deterding et al. (2011) defines gamification as integrating game elements into a secondary purpose, or more simply, the application of gaming onto another non-play construct; serious games can be considered as the heavy integration of gamification into the non-play construct. Deterding et al. (2011) present gamification as a space on a two-dimensional axis of play, and completeness (how fragmented or complete the game matches to the entire activity), described in Figure 3-6.

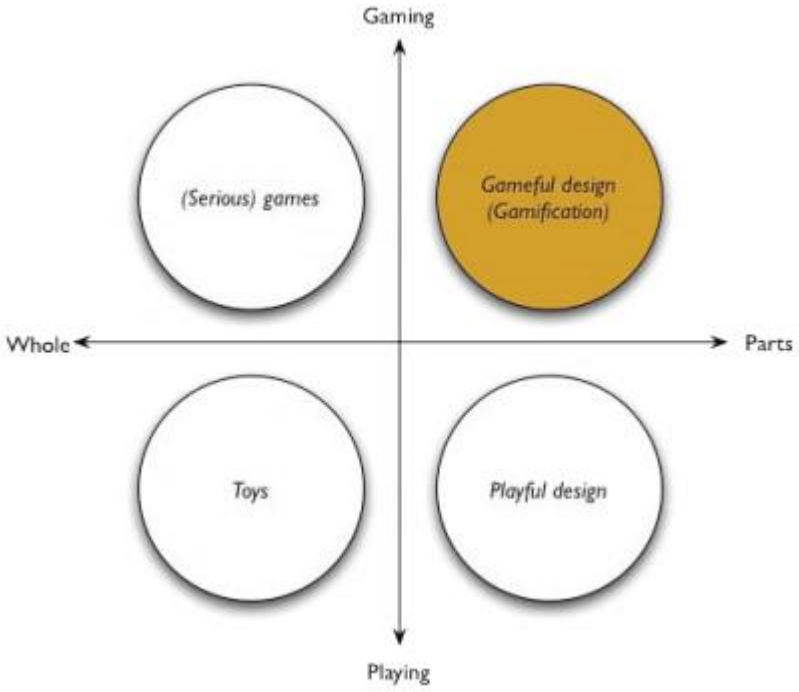


Figure 3-6 Gamification Two-Dimensional Model

While the preceding model describes gamification as separate to Serious Games, it can be noted that the completeness axis is a sliding scale, where the level of integration of gamification into the system informs the 'gamefulness' of the system; it is after a certain point that the game design informs the gamification, and not the other way around (Deterding et al. 2011). While gamification describes the conceptual idea of creating games with an incentivisation structure, there is debate as to how best to translate the desired outcome of behaviours, into game mechanics. Self-determination theory, as presented by Mekler et al. (2015) claim that empowerment of the individual user through intrinsic motivation can increase feelings of competence. This study relies on the use of points, badges and leaderboards (PBL), and demonstrated that while satisfaction increased with PBL, there was no correlation between increased performance in the activity and feelings of competence. It is apparent that gamification must aim more broadly than PBL when mapping game elements to desired outcomes. Further afield, research into MMORPGs has yielded models of mapping game elements to desired outcomes (Rapp 2017). While the authors of the aforementioned research claim a comprehensive map of game elements to desired behaviours, the analysis is far from generic, and is rooted in social gaming elements and structures. Deterding et al.

(2011) argue that gamefulness needs to take a greater precedence over playfulness in serious game design, or more precisely, focusing on the elements by which gaming occurs through, over focusing on elements of play. It is with these thoughts in mind, that a comprehensive game design framework, for use in serious game contexts is yet to become available and be fully vetted by the body of knowledge.

3.4.1.5.2 LM-GM Model

Arnab et al. (2015) describe a model of translating learning mechanics to those of game mechanics, through the use of the LM-GM model. The core component of the LM-GM is the *Serious Game Mechanic* (SGM), which is: *"the design decision that concretely realises the transition of a learning practice/goal into a mechanical element of game-play for the sole purpose of play and fun"* (Lim et al. 2016, p.1). More simply, an SGM is a core translation unit, that converts game mechanics into concrete learning mechanics, where the aggregate of these SGMs, *"support intrinsic experiential learning"* (Arnab et al. 2015). The LM-GM has been employed in a variety of contexts, from soft-skills education (Imbellone et al. 2015), to sexual education (Arnab et al. 2013), to university level engineering courses (Callaghan et al. 2016). describes the LM-GM framework visually. It can be observed that the right column of learning mechanics is derived from extant theories on learning and pedagogy and are employed in a variety of educational materials. The left column describes game mechanics, through which gameplay is enabled. The derivation of these game mechanics comes in part from the work of the Game Ontology Project (Zagal & Bruckman 2008) and the Game Object Model (Amory 2007).

These mechanics describe a wide-range of possible interactions the player of a video game can have. **Error! Reference source not found.** visually describes a possible categorisation of the LMs and GMs, into a stratified categorised model. It can be observed that LMs and GMs are stratified across six Thinking Skills, which are based on the work of Bloom et al. (1956), for describing strategies of learning. While the LM-GM model captures a wide variety of game and learning mechanics, the model is not exhaustive, and the potential for alternate or rethought mechanics is possible. The usefulness however of the tool lies in the comprehensive presentation of game

mechanics, which are a useful link between tools such as the BCW, and GOP for analysing behaviour change, and game mechanics respectively.

This study will attempt to use the LM-GM framework, for translating the outcomes of the BCW process, described in Chapter 5, to meaningful gameplay mechanics.

3.4.2 *Phase 2: Experiment Design*

The following sections will describe the tools used in this study with respect to Phase 2: Experiment Design. The ordering of the tools with respect to the experimental procedure are described in Figure 3-7.

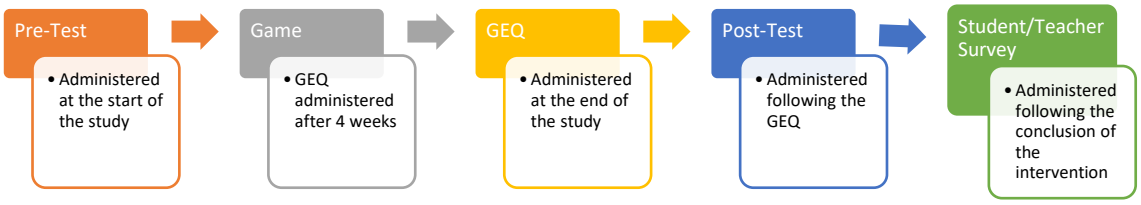


Figure 3-7 Experiment Design Procedure

3.4.2.1 Pre/Post-Tests

As the efficacy and accuracy of the assessment measures will determine the accuracy and validity of the learning outcomes within the students, the need for standardised testing is apparent (Spörer et al. 2009). As such, certain test development organisations such as Australian Curriculum Assessment and Reporting Authority (ACARA) utilise standard tests such as the *Statements of Learning for English* developed by the Curriculum Corporation (ACARA 2017e). These statements attempt to standardise the expectations for curriculum and testing in Australian schools.

The purpose of this study's pre/post-tests is to determine the state of the participant's knowledge in relation to the educational content. The tests are multiple-choice questions and answers that are of similar structures to those of the Progressive Achievement Test (PAT-R) (ACER n.d.), or the Australian NAPLAN tests in literacy (ACARA 2017a), which are calibrated and normed measures for assessing the state of a person's literacy capabilities.

Due to the standardised nature of Australian literacy education, NAPLAN testing is administered to all year 3, 5, 7, and 9 students in Australia on a yearly basis (ACARA 2017a). The standardised NAPLAN tests have created the space for many

ancillary services, which allow educational providers to access NAPLAN aligned content. Education Services Australia, a joint state owned not-for-profit provides access for educational professionals standardised assessments aligned to the Australian Curriculum, with the ability to administer tests to students electronically. The content is provided to the *Improve* service by state and federal government educational bodies and aligns with the Australian Curriculum.

The student pre-test phase of the intervention involved surveying the current state of student knowledge with respect to reading comprehension performance. A series of 18 calibrated questions, aimed to challenge the students of different reading levels. This pre-test is described in greater detail in Appendix I. The purpose of the pre-test is to provide a baseline to student reading performance, through which a post-test can act as a second baseline of student performance, following the conclusion of the intervention period.

The pre and post-test quizzes draws texts, questions, and answers from the *Improve* service and are split into three question sub-types (Australia 2017). The sub-types are author intent, inferring meaning, and identification of word meaning in context. These sub-types form an integral component of reading comprehension; the ACARA (2017b) version 7.5 lists these skills as (ACELY1698), (ACELY1702), and (ACELY1703) respectively. Both the pre-and post-tests are composed of 18 question and answer multiple-choice quizzes that resolve to four metrics: inference score, author meaning score, word meaning score, and the total reading comprehension score. The scores provide a metric of comparison for the four previously mentioned factors. The test can be viewed in full in Appendix I and Appendix II.

3.4.2.2 Game Experience Questionnaire

While significance of learning outcomes can be measured quantitatively, it is important to determine how well the game system fulfilled student requirements. Serious games have the dual purpose of providing instruction, while also aiming to be entertaining. The entertainment value of the system is therefore as vital as the magnitude of change that each student may experience insomuch as continual participation with the system is well supported by meaningful play as described previously in Section 2.2.1. This study utilises the Game Experience Questionnaire (GEQ), designed by IJsselsteijn

et al. (2013) and provides a normalised pattern for describing participant experiences following a video game experience. The GEQ resolves to several normalised scores which indicate an overall picture of the following factors relating to video game experiences. The GEQ is divided into four separate iterations of the quiz: a core test, a social test, an in-game test, and a post-game test. This study makes use of the post-game test for two reasons. Firstly, the GEQ was administered twice over the course of the study, at the end of a play session. Secondly, fatigue with students was a primary concern, with respect to question and answer surveys. The post-game module provides a reduced question set, without compromising the validity of the measure. The reading age of this test was confirmed as approximately Grade 2 (age 7– 8), when assessed using an averaged reading age composite checker (readable.io 2017).

Questions	**GEQ Core**	**Components**
1. I was interested in the game's story	GEQ Core 3	Sensory and Imaginative Immersion
2. I felt successful	GEQ Core 17	Competence
3. I felt bored	GEQ Core 16	Negative Affect
4. I found it impressive	GEQ Core 27	Sensory and Imaginative Immersion
5. I forgot everything around me	GEQ Core 13	Flow
6. I felt frustrated	GEQ Core 29	Tension
7. I found it tiresome	GEQ Core 9	Negative Affect
8. I felt irritable	GEQ Core 24	Tension
9. I felt skilful	GEQ Core 2	Competence
10. I felt completely absorbed	GEQ Core 5	Flow
11. I felt content	GEQ Core 1	Positive Affect
12. I felt challenged	GEQ Core 26	Challenge

13. I had to put a lot of effort into it	GEQ Core 33	Challenge
14. I felt good	GEQ Core 13	Positive Affect

Table 7 GEQ Post-Game Survey

As displayed in Table 7, the GEQ Post-Game Module breaks game experience down into four categories of experience. The GEQ was administered twice during this study. On the first occasion, half-way through the intervention period at 4 weeks, and again at the conclusion of the study at 8 weeks. The purpose of multiple GEQ tests is to determine whether gaming experience changes over time, across cohort and within the same student.

3.4.2.3 Student Feedback

The student feedback survey, as previously mention in aims to gather subjective, and demographic information on the students. The purpose of this feedback is to determine whether demographic subsets of the cohort have meaningful interactions with the system, in this manner, SRQ4 can be answered. Along with the demographic survey, opinions on the system were gathered, to provide information on the user experience of the system. The survey is described fully in Appendix III

3.5 Summary

This Chapter presented the core methodology used throughout this study. Tools for the design of a serious game intervention were described. The Behaviour Change Wheel methodology was detailed, along with the LM-GM framework for mapping game mechanics to learning mechanics. Tools for collecting data to answer the research questions were also presented, including a description of the pre/post-tests, and the Game Experience Questionnaire. The following Chapter will detail the way the serious game used in this study was developed, and the way it collects, stores, and processes the information needs presented in Section 3.2.

4 System Design

As identified in the preceding Chapters, attention and motivation are key components into the success or failure of serious game solutions for children. It is therefore imperative that both behaviours are fully examined in the context of a behavioural framework, such that a rigorous process of game design may be overlayed upon the said examinations and findings. In simpler terms, it is important to fully understand the problem, before a solution is designed for it.

4.1 *Phase 1: System Design*

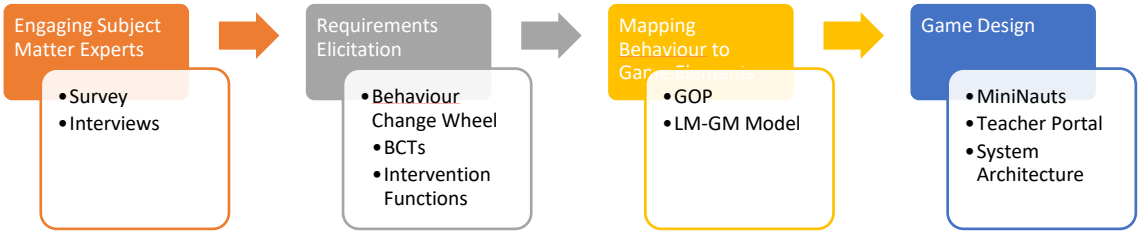


Figure 4-1 System Design Process

As described in the previous chapter, Figure 4-1 indicates the tools that were employed throughout the design phase of this research project. The following sections will now describe the application of each of these tools.

4.1.1 *School Context*

Educational settings are increasingly monitored and accountable for the performance of each student (ACARA 2017b). In this climate, teachers are under pressure to deliver standardised content to their students, and to achieve the highest test scores possible. Funding for the school, and even the teacher's employment may be tied to these test scores. It is therefore imperative that any Serious Game System compliments the current standards of their classroom, and can deliver meaningful results in student's learning, as well as in their test scores. This research engaged with two teachers to determine what the desirable form and outcomes of reading comprehension video game would be. Through interviews with two public primary school teachers, a list of requirements was derived. These interviews took place over several weeks and were

through the mediums of face-to-face conversations, and emails. Teachers were presented with a series of open-ended questions and asked for their opinions on a variety of topics, including game design, student motivation, and their opinions on barriers to literacy education. This survey is presented in Appendix IV.

The game must run itself, and not require regular input from the teacher

Teacher participants engaged in this study expressed a desire for a serious game system which would to many extents be as autonomous as possible. In the words of one teacher: "*We don't have the time to play around with things, it really just needs to do what we tell it [sic] do, and then off it goes.*" This desire for an autonomous system fits within the literature regarding serious games (Susi et al. 2007), and even more broadly technological literacy intervention As previously described in Section 2.1.4.3, teachers have a highly limited time in the classroom per student, and any task that could be automated to provide students with feedback on progress, as well as simply administering tasks, would be of benefit.

The game must be fun and engaging for students, but must be educational

Teachers felt the need to elaborate that while their students would enjoy playing video games during class time, students needed to be presented with meaningful educational opportunities every time they interacted with the game. The time of teachers to educate students is a limited resource, and time spent in play during class time must be mediated by some educational benefit. This was a requirement placed on them by the school, and by the state education body; *"We only have a limited time in class for these sorts of things, it's gotta be quick, and easy."*

The game should provide feedback for the teacher on student and class performance as a whole

A core motivating driver for this research is the continued expressed desire by the teacher participants, for the serious game solution to include comprehensive feedback mechanisms. These mechanisms, as identified by the teacher needed to include feedback for the students, illuminating their current progress, how they have progressed over time, and where they can improve. Secondly, teachers required a system to effectively administer content and determine the scores and comprehension of

individual students. Specific mechanisms for feedback to the teacher are presented in Section 4.7.4 and feedback mechanisms for students are detailed in Section 4.5.6.

The game should be re-usable and not lose novelty quickly

It was identified by the teacher participants in this research that any serious game solutions employed in their classroom would need to sustain a measure of longevity. As previously discussed in Section 2.2.2.6, the concept of *content agnostic* serious games was presented; serious games which could act as a framework for educational content, but do not deeply embed this content into the game. Teachers identified that any serious game solution would need to work for at least one term, and beyond, as student interest in the system was paramount. Teacher participants felt that if the serious game was too repetitive, or students had to engage in the same narrative over and over, there would be a distinct lack of interest. This requirement was enacted and enabled through the use of mini-games (refer to Section 4.5).

4.1.2 *Requirement Discussion*

Through the requirements elicitation process, it was determined that teachers have a wide range of opinions with respect to serious games. Both teachers expressed a positive attitude toward serious games as a concept, and had anecdotally used various serious games such as *Mathletics* (Learning 2017) in their classrooms to their satisfaction. Teachers were also enthusiastic about the concept of a reading comprehension serious game, as in their experience, they had not encountered any on the internet which were satisfactory for a later primary school market. The positive feelings towards serious games were mediated however with several reservations. Teachers were wary of change in their classroom, and any new educational artefact should be measured on its usefulness to education and its ability to hold student's attention. It is with these points in mind, that the game system was designed as a mechanism to delivering meaningful reading experience for students.

While the system has been described in terms of requirements, it is necessary to convert the behavioural modelling of the Behaviour Change Wheel (BCW) to video game elements, using a standardised framework. This list of requirements, was used as

the starting requirements and conditions for the Behaviour Change Wheel analysis of student behaviour, see Section 4.2.

4.2 *Application of the Behaviour Change Wheel to Elicit Requirements*

4.2.1 *Understand the Behaviour*

As previously described, the Behaviour Change Wheel is a tool and more importantly a process for designing interventions. Section 4.2.1.1, Section 4.2.1.2, Section 4.2.1.3, and Section 4.2.1.4 describe Stage 1 of the BCW process, as detailed in Figure 4-2.

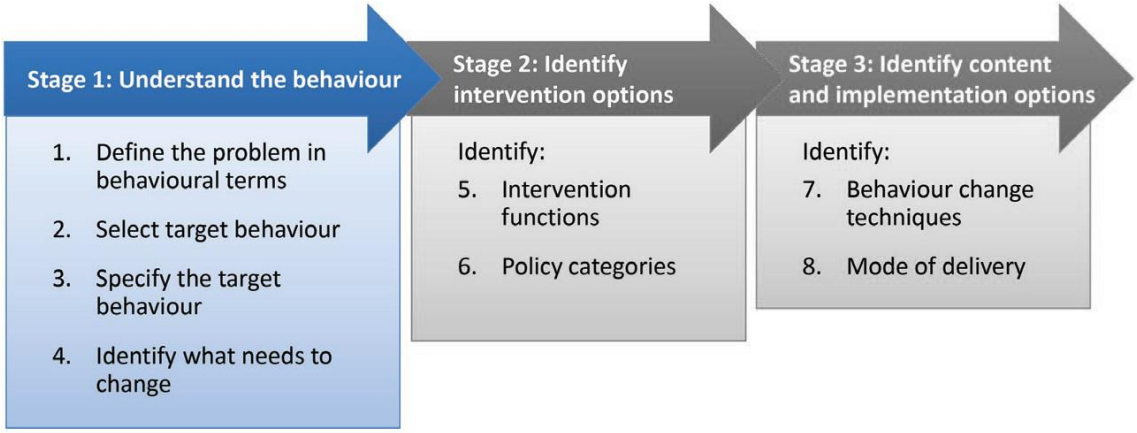


Figure 4-2 BCW Process Stage 1—Taken from Michie et al. (2014)

4.2.1.1 Define the Problem in Behavioural Terms

Students frequently disengage with reading materials, in both school and home settings Machet (2002). The problem may arise due to a lack of interesting content, the stigma of reading, or the normalisation of reading habits (Richmond et al. 2008). Requirement elicitation came in the form of two, one-hour interviews held with the teacher participants of this study. Teachers were asked a broad range of questions, with a focus upon the following points:

- Identifying the single greatest difficulty, with respect to literacy education, that faces students on a personal learning level.
    - Reading comprehension was indicated as the skill that was lacking the most in the classroom.

  - Reading comprehension was also stated to be a core skill that informed how well a student would absorb literacy activities, and indeed how well they generally interacted with learning in class activities.
  - The deficit in reading comprehension was considered to be severe for several students, and moderately concerning for a small portion of the classes, particularly for males.
- How primary school teachers administers literacy course material, and how this relates to the Australian Curriculum.
  - Teachers indicated their general teaching process as follows:
    - Discuss a topic broadly with the class.
    - Ask class to perform a class-wide activity.
    - Break the class into groups to perform further literacy activities
    - Allow students individual reading time, with testing of their reading occurring either with flash cards from their peers, or by teachers choosing the student to state their answer to a question with the class.
- How the teacher identifies student progress toward satisfactory requirements.
  - Teachers discussed how their own internal scoring was not comprehensive, and many students were identified as being poor in the reading comprehension, but measurements to their overall improvement were lacking
    - This lack of measurable improvement was attributed to the lack of time teachers could spend individually with students
  - Yearly testing provides teachers with an opportunity to view how well their class performs nationally, with respect to literacy via NAPLAN testing.
    - The results are delayed by up to six months.
    - Only for year 5 students.
    - Are not segmented by class, only by school.
- In what ways are students tested for the knowledge of literacy?
  - Testing opportunities, as previously stated is few and far between

        - Several times a year, teachers run the PAT-R test in their classes to measure changes in performance
- In what ways, could a game system be utilised in a classroom environment by teachers, and students?
    - Teachers identified that a game system could be highly beneficial to student literacy.
        - Male students were identified as being particularly enthusiastic about computer time in class, especially for playing games.
        - Game systems which could automate the process of assessing literacy tasks would be of great benefit to the teacher.

The identified target behaviour for the behaviour change wheel was identified by the researcher, through this requirement gathering process as: *Increasing the motivation for students to engage with reading comprehension activities*. This behaviour occurs mostly in the school environment. The identified problem concerns voluntary reading, but the lack of ability to read complex sentences can spill over into day-today reading activities, which occurs in the home and community. The identified cohort are children, in the age range of 10–12. This age group is especially prone to problems with reading comprehension, as societal pressures and accumulate reading deficits may begin to weigh heavy on students, forcing them to disengage. This information is summarised in Table 8.

Questions	**Responses**
What behaviour?	Low motivation and performance in reading comprehension activities
Where does the behaviour occur?	School, Home, The Community
Who is involved in performing the behaviour?	School aged children aged 10–12

Table 8 Identified Behaviour

The following sections deductively follow a process of identifying an axiom, and elaborating further upon it, until it can be summarised succinctly, and be used in a later stage of the BCW. This approach to reasoning out a desired outcome from a targeted behaviour is further elaborated in Michie et al. (2011).

4.2.1.2 Select the Target Behaviour

Selecting Target Behaviours involves identifying solutions to the problem behaviour, by mapping what the correct behaviour may look like. The following list presents potential target behaviours, following discussions with two teachers in a Tasmanian school, along with extant literature provided in Chapter 2, particularly Section 2.1.4. The list of the following correct behaviours has been generated specifically from teacher interviews prior to the commencement of the study, which asked the teachers what areas they believed were the sources of trouble for literacy, and how these areas would look if the behaviours were correct. This stakeholder consultation follows the recommendation of Michie et al. (2014).

1. Engage with parents to read more often at home.
2. Engage with the teacher for more face-to-face time, to aid the student
3. Reflect on reading comprehension activities undertaken, by themselves, or with others
4. Regularly practice reading comprehension activities in the classroom, at specified intervals
5. Read more novels or stories, to practice reading comprehension skills.
6. Allow students to choose their own reading material to engage with in class
7. Engage with non-traditional reading comprehension activities (non-paper) (Prensky 2001, Hamari et al. 2014, Segers & Verhoeven 2002, Dede 2005)

Target behaviours, according to Mitchie (2012), need to be ranked, to determine the efficacy of each behaviour based on the following criteria:

1. How much of an impact changing the behaviour will have on desired outcome?
2. How likely it is that the behaviour can be changed (when considering likelihood of change being achieved, think about the capability, opportunity and motivation to change of those performing the behaviour)?

3. How likely it is that the behaviour (or group of behaviours) will have a positive or negative impact on other, related behaviours?
4. How easy it will be to measure the behaviour?
5. How much will the behaviour change the person's situation toward the desired outcome?
6. How possible is it for the behaviour to be changed, in a meaningful way?
7. How will the behaviour affect other behaviours in the individual, in positive and negative terms?
8. How measurable is the target behaviour?

Each behaviour through prioritisation, can then be categorised into one of the following categories:

- The behaviour appears very promising as a target behaviour (■).
- The behaviour is quite promising as a target behaviour (■).
- The behaviour appears unpromising but is worth considering as a target behaviour (■).
- The behaviour is not acceptable as the target behaviour (it doesn't matter what it is like on the other criteria, this behaviour cannot be selected as the intervention target) (■).

The identified target behaviours relate highly to teacher-student interaction. While home environments can be used to support the desired outcome, schools act as a vital focal point for students to explicitly practice their comprehension skills. The identified target behaviours should be few (one-three). The following table describes the prioritisation process and has been generated regarding extant literature, and stakeholder consultation.

Potential target behaviours	**Impact of behaviour change**	**Likelihood of changing behaviour**	**Spillover score**	**Measurement score**

Engage with parents to read more often at home.	Unpromising but worth considering	Unpromising but worth considering	Promising	Unpromising but worth considering
Engage with the teacher for more face-to-face time, to aid the student	Very Promising	Promising	Very Promising	Promising
Reflect on reading comprehension activities undertaken, by themselves, or with others	Promising	Promising	Unpromising but worth considering	Unpromising but worth considering
Regularly practice reading comprehension activities in the classroom, at specified intervals	Very Promising	Promising	Very Promising	Promising.
Read more novels or stories, to practice reading comprehension skills.	Promising	Promising.	Promising.	Unpromising but worth considering
Allow students to choose their own reading material	Promising.	Unpromising but worth considering	Unpromising but worth considering	Unpromising but worth considering

to engage with in class				
Engage with non-traditional reading comprehension activities (non-paper)	Promising	Promising	Very Promising	Very Promising
Selected Target Behaviour	**1. Engage with the teacher for more face-to-face time, to aid the student** **2. Regularly practice reading comprehension activities in the classroom, at specified intervals** **3. Engage with non-traditional reading comprehension activities (non-paper)**			

Table 9 Prioritisation of Potential Target Behaviours

Through the prioritisation process presented in Table 9, three potential target behaviours were identified. These three behaviours, once defined, can now be expanded upon in detail.

4.2.1.3 Specify the Target Behaviour

Target behaviours need to be clearly defined, to analyse their benefit toward the outcome behaviour. This definition process is defined in Table 6 (APEASE Criteria Definitions), Table 10, Table 11 and Table 12.

Target Behaviour	**Engage with the teacher for more face-to-face time, to aid the student**
Who needs to perform the behaviour?	Teachers and students
What do they need to do differently to achieve the desired change?	Teachers need to spend more time individually observing and interacting with each student. Each student needs to wait for the teacher to interact with them during class.
When do they need to do it?	During reading comprehension activity time.
Where do they need to do it?	At school.
How often do they need to do it?	During specified class times for reading comprehension, according to teacher plans.
With whom do they need to do it?	Teachers and students.

Table 10 Target Behaviour One Analysis

Target Behaviour	**Regularly practice reading comprehension activities in the classroom, at specified intervals**
Who needs to perform the behaviour?	Students in the classroom.
What do they need to do differently to achieve the desired change?	Students need to work individually or in a group to practice their reading comprehension skills through measurable and repeatable activities. Students need to focus on the task and perform the necessary reading skills to infer meaning from text.
When do they need to do it?	During reading comprehension activity time.
Where do they need to do it?	At school.
How often do they need to do it?	During specified class times for reading comprehension, according to teacher plans. Outside of regular class time.
With whom do they need to do it?	Not dependant on others.

Table 11 Target Behaviour Two Analysis

Target behaviour	**Engage with non-traditional reading comprehension activities (non-paper)**
Who needs to perform the behaviour?	Students in classrooms.
What do they need to do differently to achieve the desired change?	Engage with different technologies/media that present reading comprehension activities.
When do they need to do it?	During reading comprehension activity time. Outside of regular class time.
Where do they need to do it?	At school.
How often do they need to do it?	During specified class times for reading comprehension, according to teacher plans. Outside of class (at home).
With whom do they need to do it?	Not dependent on others.

Table 12 Target Behaviour Three Analysis

This analysis process yielded several interesting results. Firstly, it was identified that engagement with reading comprehension material was of a high importance to the *what* component of the analysis. The repeated nature of this practise too surfaced as a consideration in the *how often*. While the dependency of students upon their teacher is made apparent in the first target behaviour, the core component of teacher interaction with students is to perform a validating and feedback mechanism for student reading work. Once these behaviours have been identified, the core *what* of the change management process needs to be identified. This is presented in the following section.

Now that these behaviours have been identified, the core *what* of the change management process needs to be identified. This is presented in the following section.

4.2.1.4 Identify What Needs to Change

Identifying what needs to change requires using the Capability, Opportunity, Motivation Behaviour (COM-B) model to understand how to facilitate the new target behaviours. As previously described, the COM-B model provides insight into the exact mechanisms and requirements of the behaviour change process, before the intervention

can be designed. Due to the similarities of the first and second target behaviours, the two behaviours are analysed together in terms of needs (as shown in). To restate, these target behaviours are:

1. *Engage with the teacher for more face-to-face time, to aid the student,*
2. *Regularly practice reading comprehension activities in the classroom, at specified intervals*

COM-B Components	**Criteria**	**Responses**
	What needs to happen for the target behaviour to occur?	Is there a need for change?
Physical capability	Be physically present in the classroom with the teacher.	No change needed.
Psychological capability	Wait for the teacher to engage with the student. Perform the activity independently.	Patience and self-reliance in students.
Physical opportunity	Be present in class.	No change needed.
Social opportunity	Engage with activities that are not considered boring or stigmatising. Interact with the teacher in meaningful ways.	Yes, activities need to engage student's interest, while also being available for the teacher to work through with the student.

Reflective motivation	Believe that the teacher can help them. Believe that they can perform the work independently.	Students need to engage in activities and then understand what occurred, how the received the mark they did.
Automatic motivation	Be willing to engage with the activity in the class, such that motivation becomes an intrinsic factor.	Activities need to be engaging, and feedback needs to be relevant for the student, such that engagement is not reliant on the intrinsic fun of the activity.
Behavioural diagnosis of the relevant COM-B components:	• **Automatic Motivation** • **Reflective Motivation**	

Table 13 Target Behaviour One & Two Needs

The third identified Target Behaviour, as previously listed in Section 4.2.1.3, was determined to be:

3. *Engage with non-traditional reading comprehension activities (non-paper).*

Table 14 Target Behaviour Three Needs details the process of analysing the Target Behaviour with respect to COM-B components.

COM-B Components	**Criteria**	**Responses**
	What needs to happen for the target behaviour to occur?	Is there a need for change?
Physical capability	Be physically present in the classroom with the teacher. Be capable of engaging in non-paper mediums.	No change needed.

Psychological capability	Perform the activity independently. Comprehend the new activity, and how it relates to the real world.	Patience and self-reliance in students.
Physical opportunity	Be present in class.	No change needed.
Social opportunity	Engage with activities that are not considered boring or stigmatising.	Activities need to be stimulating and novel. Activities need to be presented with a low barrier for entry for students.
Reflective motivation	Believe that the activity is relevant for their reading, and that it can be fun. Be provided with feedback opportunities that model the teacher-student process.	Students need to engage in activities and then understand what occurred, how the received the mark they did.
Automatic motivation	Be willing to engage in a non-paper-based activity.	No change needed.
Behavioural diagnosis of the relevant COM-B components:	• Automatic Motivation • Reflective Motivation	

Table 14 Target Behaviour Three Needs

Through the needs analysis presented in Table 13 and Table 14, it was determined that all three target behaviours resolved to the Automatic and Reflective Motivation COM-B components. To generalise the process, it was identified that the major need for the target behaviours to be implemented is the lack of motivation that may be present in current reading comprehension activities. The next phase of the BCW

process identifies manners in which the target behaviours can be functionally implemented.

4.2.2 *Identify Intervention Options*

Section 4.2.2.1 are the application of the BCW Stage 2 process of identifying intervention options (refer to Figure 4-3). Note that *Step 6. Policy Categories* has been removed from this process, due to the constraints of ethical approval in Tasmanian schools prescribing certain policy procedures upon this research, particularly those listed in Section 5.5; policy was pre-determined to include a serious game for change administered by teachers in Tasmanian classrooms

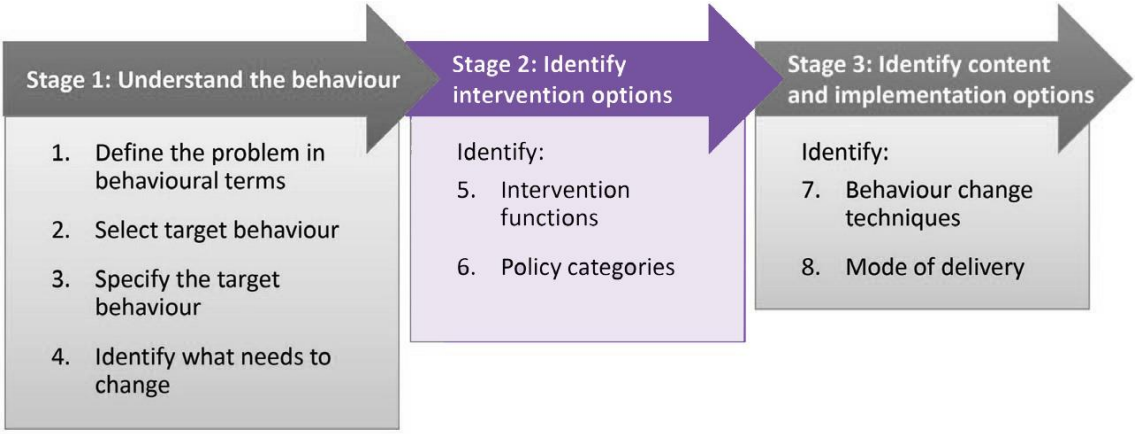


Figure 4-3 BCW Process Stage 2—Taken from Michie et al. (2014)

4.2.2.1 Identify Intervention Functions

Once the target behaviours are analysed for their need for change, against the COM-B model, the behaviours become mapped to intervention functions, which can be reviewed in Michie et al. (2014). Intervention functions are analysed for their suitability via the APEASE criteria. The APEASE criteria (as previously listed and explained in Table 6) are as follows:

- Affordability;
- Practicability;
- Effectiveness and cost-effectiveness;
- Acceptability;
- Side-effects/safety; and
- Equity.

The APEASE criteria are applied to the three identified target behaviours in Table 15 and Table 16. While the APEASE criteria demonstrate the metric by which target behaviours should be assessed, it should be noted that a discretionary approach is undertaken by the reviewer, using the guide of extant literature, along with personal experience of themselves and stakeholders, to assess the manner in which target behaviours should be assessed and ranked. This follows the assertion by that intervention function design should be part of an expert lead approach (Michie et al. 2014)

Target Behaviours

1. **Engage with the teacher for more face-to-face time, to aid the student**
2. **Regularly practice reading comprehension activities in the classroom, at specified intervals**

Candidate intervention functions	**Suitability**	**Does the intervention function meet the APEASE criteria (affordability, practicability, effectiveness/cost-effectiveness, acceptability, side-effects/safety, equity)?**
Education	Suitable	Students and teachers can be educated as to the benefit of greater face-time with teachers, along with regular intendent practice.
Persuasion	Not Suitable	Persuasion is a regular component of school environments. However, due to the problems surrounding motivating non-compliant or low-compliant students with reading, persuasion may not act as an enabler.
Incentivisation	Suitable	Incentivising students to engage in independent activities, while being monitored by a teacher can have an impact on the desired outcome.

Coercion	Not Suitable	Coercion is no longer a supported mode for encouraging students to engage with educational materials.
Training	Not Suitable	Training is not required to impart any new abilities or skills.
Restriction	Not Suitable	Restriction refers to legal and regulatory mechanisms that are not accessible in this research study.
Environmental Restructuring	Not Suitable	Restructuring the environment is not practical within the confines of this study.
Modelling	Not Suitable	Modelling is not necessary for this study. Student behaviours have already been performed in relevant literature.
Enablement	Suitable	Enablement is possible to allow the student to engage with activities independently.
Selected Intervention Functions		**Education** **Incentivisation** **Enablement**

Table 15 Target Behaviour One & Two APEASE Analysis

Target Behaviours **Engage with non-traditional reading comprehension activities (non-paper)**		
Candidate intervention functions	**Suitability**	**Does the intervention function meet the APEASE criteria (affordability, practicability, effectiveness/cost-effectiveness, acceptability, side effects/safety, equity)?**
Education	Suitable	Students and teachers can be educated as to the benefit of interacting with new media.

Persuasion	Not Suitable	Persuasion is a regular component of school environments. However, due to the problems surrounding motivating non-compliant or low-compliant students with reading, persuasion may not act as an enabler.
Incentivisation	Suitable	Incentivising students to engage in independent activities using different media is possible. The incentivisation can be intrinsic to the activity, or as a reward for the behaviour.
Coercion	Not Suitable	Coercion is no longer a supported mode for encouraging students to engage with educational materials.
Training		No Training is not required to impart any new abilities or skills.
Restriction	Not Suitable	Restriction refers to legal and regulatory mechanisms that are not accessible in this research study.
Environmental Restructuring	Not Suitable	Restructuring the environment is not practical within the confines of this study.
Modelling	Not Suitable	Modelling is not necessary for this study. Student behaviours have already been performed in relevant literature.
Enablement	Suitable	Yes Enablement is possible to allow the student to engage with activities independently, particularly new media.

Selected Intervention Functions	**Education**
	Incentivisation
	Enablement

Table 16 Target Behaviour Three APEASE Analysis

Through mapping intervention functions, it was determined that Incentivisation, Education, and Enablement were the most promising functions to apply to the three identified target behaviours.

4.2.2.2 Policy Categories

As previously explained in Section 4.2.2, this research is constrained by ethical standards guiding the interaction and design of interventions with children in Australia. Prior to serious modelling, ethical approval through university and state bodies was sought, and this research was approved following the guidelines outlined in Section 5.5.

4.2.3 *Identify Content and Implementation Options*

Section 4.2.3.1 and 4.2.3.1 describes the application of Stage 3 of the BCW methodology for this study (refer to Figure 4-4).

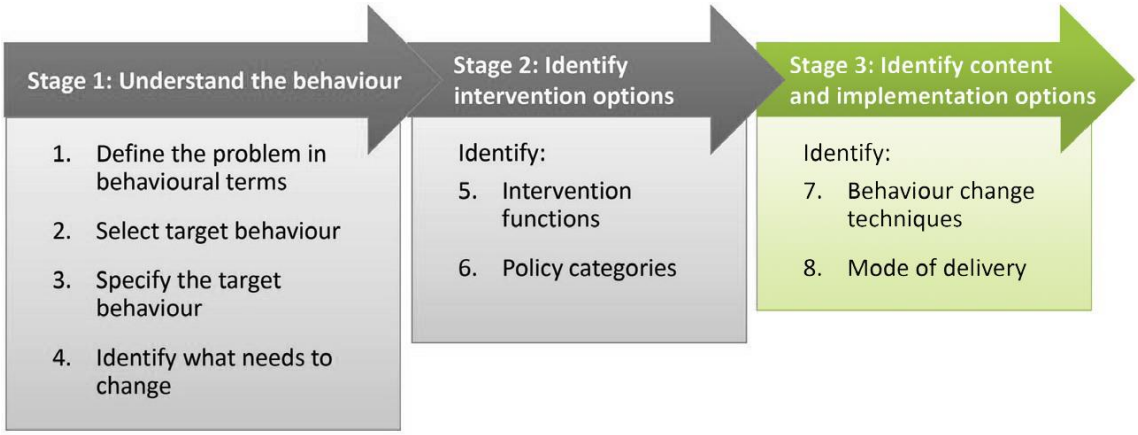


Figure 4-4 BCW Process Stage 3—Taken from Michie et al. (2014)

4.2.3.1 Identify Behaviour Change Techniques

The seventh stage of the BCW intervention design is to identify behaviour change techniques which correspond to intervention function elements (Michie et al. 2013). Through the preceding stage, it was determined that using non-traditional reading comprehension activities is a valid Intervention Function, when analysed via

the APEASE criteria. Relevant BCTs are defined by Michie et al. (2013) as modes for enabling the desired behaviour.

BCT	Description
Feedback on behaviour	Provide opportunities to describe current behaviour, and how it could be improved
Identity associated with changed behaviour	Create a sense of ownership over the target behaviour, providing a space for a person to own their new behaviour
Information about others' approval	Provide opportunities for approval of a person's new behaviour to be given
Identification of self as role model	Use the individual, and their self as a benchmark for current, past, and future sense of progression toward the target behaviour
Salience of consequences	Provide opportunities for the person to be reminded of the outcome of their actions
Incentive	Provide motivation to the person through the use of motivating external and internal items
Discrepancy between current behaviour and goal	Demonstrate the magnitude of the difference between the current state of the person's behaviour, and where they could be
Punishment	Deny opportunities for rewards, approval, or commendation if the target behaviour is not worked toward
Behaviour cost	Related to *Punishment*, demonstrating the magnitude of a person's actions

Table 17 Identified BCTs

As shown in Table 17, a list of nine BCTs relevant to the identified Intervention Functions of Incentivisation, and Enablement were chosen. As discussed in Chapter 2,

literature on Incentivisation, and Enablement point to a variety of game mechanics through which these functions can be enabled (Garris et al. 2002; Abrams & Gerber 2013; Huang et al. 2010; Amory & Naicker 1999). Section 4.3 describes the translation of the BCTS to game mechanics, and the manner in which this is supported by the body of knowledge.

4.2.3.2 Identify Modes of Delivery

The BCW defines a process whereby the mode of delivery can be identified and measured for its efficacy against the APEASE criteria. Through analysing factors such as distance, cost and equity, it was determined that technological solution would be most appropriate. This study proposes using a medium such as video games to enable students to engage in independent learning. Alongside this, students can be incentivised with video games as an expression of play while also engaging in the desired behaviour. Most importantly, video games provide excellent opportunities for independent reflection on feedback by students, as discussed in Section 2.2.1.4. The selection of this mode of delivery was deemed suitable and detailed in Chapter 2. Video games are at the core, feedback machines, and any interaction with the system, involves experimentation, enjoyment and feedback, all desired facets of enablement and incentivisation.

4.3 Application of the LM-GM Analysis to Map Game Elements to Behaviour

The outcome of the BCW analysis process, described in Section 4.2.3.1 was a list of identified BCTs (see Table 17), which support identified Intervention Functions (IFs). The selection of the three IFs presented in Section 4.2.2.1 formed the starting point for the selection of relevant BCTs. Whilst the BCW provides a process-driven methodology for generating information, and strategies relating to behaviour change, the mechanisms through which these strategies can be implemented has so far not been discussed. Research into effective translation between behaviour change methodologies, and serious game design are few and far between. Research into providing taxonomies on game mechanics provide a starting point for matching BCTs to game mechanics. The

Game Ontology Project (GOP) (Game Ontology Project 2017) seeks to provide the research community with a comprehensive framework for analysing, and describing video games (Game Ontology Project 2017). The GOP has been described by Dormans & others (2012) as a vocabulary for understanding video games, and their constituent elements. Whilst the GOP does not provide a direct framework for designing serious games from behavioural principles, it does provide the opportunity for describing game mechanics that fit to behaviour change techniques (Zagal & Bruckman 2008).

Game Elements listed in Table 18 are generated from the Game Ontology Project's comprehensive list of game elements and mechanics (Game Ontology Project 2017). Elements were chosen from the GOP for their relevance and similarity to the BCTs described by (Michie et al. 2013), which were selected in Section 4.2.3.1. The game elements listed form a starting point for designing a serious game, while maintaining a strong basis in extant behavioural practice.

BCT	**GOP Game Elements**	**Description**
Feedback on behaviour	Feedback	Player is rewarded through game progress and score if the desired behaviour occurs.
Identity associated with changed behaviour	Avatar/Game Customisation	Identity is associated with the customisation of an in-game character
Information about others' approval	Feedback	The game provides elaborative and supportive feedback when correct/incorrect behaviour occurs
Identification of self as role model	Performance Record	Show players where they could improve with their result.
Salience of consequences	Goal Metrics/Game Goals	Behaviour in game affects student performance in the game; progress in game is linked to reading comprehension skill

Incentive	Rewards	Rewards are provided to the player on completion of goals
	Rewards	Show where rewards were gained and where certain activities could have enabled more rewards
Discrepancy between current behaviour and goal	Performance Record/Required Goals	Success in game goals linked to correct reading comprehension activity; Poor performance in reading comprehension means that students do not perform well at the game
Punishment	Rewards	Rewards are not forthcoming if students don't complete their activities
Behaviour cost	Rewards	Players can see that they do not receive rewards if they do not engage

Table 18 BCT and GOP Comparison

Following the identification of relevant game mechanics through the GOP, a link between behavioural techniques and potential game mechanics has been established. However, due to the general nature of the GOP, specific traits of serious game design are not yet factored in. More specifically, successful learning mechanics as described by Breuer & Bente (2010) can be placed on a spectrum of relevance to learning goals. Ritterfeld and Weber cited in Breuer & Bente (2010) propose that for a serious game to be most effective at learning, and evocative for fun, the acknowledgement that entertainment must be constrained and mediated through carefully chosen learning mechanics. Learning Mechanics (LMs) as defined by Callaghan et al. (2016), are objects and methods of instruction and enablement through which learning can be facilitated. It is at this point that a structured link between identified game mechanics, behaviour change techniques and learning mechanics can be made.

It was identified through the BCW process that Enablement and Incentivisation formed core functions for potentially successful behaviour change. As has been demonstrated in Chapter 2, serious games can enable, and incentivise players generally,

and in the context of learning. The link therefore between behaviour change, serious games, and learning therefore must be explored.

The following game and learning mechanics, listed in Table 19 LM-GM Comparison

detail the breakdown of game mechanics selected for the system, how they correspond to learning mechanics, and the way they are implemented. Rows are denoted pertaining to the identified intervention functions of the BCW, presented in Section 4.2.2 Enablement (■), and for Incentivisation (■). Numbers presented in Table 19 LM-GM Comparison

represent the identified BCTs generated from Section 4.2.3.1. It can be observed that learning and game mechanics map ideally to BCTs. This is due to the BCTs presenting an ideal technique to derive a desired behaviour, where a corresponding LM-GM mechanic serve as the action to this technique. (Arnab et al. 2013) describe the LM-GM's application as to evaluate a given game's efficacy in the classroom, after it has already been designed. The application of the LM-GM in this study takes the process further, toward designing the game from a set of behavioural requirements, where teacher input is filtered through the lens of problem and ideal behaviours. This mapping of game mechanics to behavioural techniques is novel but does sit with well with the BCW's ideal of intervention functions tailored to the individual or group setting, where impact will be most strongly felt (Michie 2012).

BCTs	Game Mechanics (LM-GM)	Learning Mechanics (LM-GM)	Implementation in Serious Game	Sec.
Feedback on behaviour [1]	Status [2,4,6,9]	Ownership	Playable/upgradeable characters	Section 4.3.1
Identity associated with changed behaviour [2]	Rewards/ Penalties [5,6,7,8,9]	Incentive	Score, points, unlocks, character level.	Section 4.3.2
Information about others' approval [3]	Feedback [1,8,9]	Feedback	End of game feedback screen	Section 4.3.3
Identification of self as role model [4]	Progression [1,3,4,6]	Action Task	Character level, score	Section 4.3.4
Salience of consequences [5]	Tutorial [7]	Tutorial	Tutorial, game help text	Section 4.3.5
Incentive [6]	Cut scenes/ Story [6]	Discover	In game cut scenes	Section 4.3.6
Discrepancy between current behaviour and goal [7]	Tokens [6]	Repetition	Unlock tokens	Section 4.3.7
Punishment [8]	Behavioural Momentum [5,7]	Repetition	Score and achievement over time	Section 4.3.8
Behaviour cost [9]				

Table 19 LM-GM Comparison

As can be seen in Table 19, LMs can be readily mapped and related to GMs, with reference to BCTs. Through this analysis process, it was identified that learning mechanics such as Feedback, Action/Task, Ownership, and Incentive strongly couple, and overlap with several BCTs. Mechanics such as Repetition, Discover, and Tutorial

rely on fewer BCTs, but remain coupled with Salience, Incentivisation, and Discrepancy through their similarity in aims. The LM-GM model proposes that knowledge acquisition must be obtained through game mechanics, over direct instruction (Callaghan et al. 2016). The following sections outline the identified game mechanics and how they relate to the identified LM-GM learning mechanics.

4.3.1 *Status*

Status, as described in Table 20, relates to the *Ownership* learning mechanic (Callaghan et al. 2016). The purpose of status is to reflect player achievement in game, to a tangible reward. It was identified that in the context of a serious game, playable characters present an opportunity to express status (Charsky 2010). If a player is allowed to express their individual performance in-game, this becomes an illustration of success, and adds to the sense of achievement of this system (Charsky 2010).

4.3.2 *Rewards/Penalties*

As previously described in Chapter 2, serious games typically revolve around a feedback loop, where correct actions are rewarded, and incorrect actions are punished in some form (Abrams & Gerber 2013). The exact mechanism of these rewards and punishments are at the discretion of the game designer. What is important, as exhorted by Gee (2005), is to provide a context in which players are empowered to play a game, and achieve the dual satisfaction of learning, while playing. The reward mechanisms identified that would suit the requirements of this system include points, and tokens, for use in identifying progress and spending on unlocks respectively. These systems sit above the core gameplay, and are transient to each individual game session, but form an important part of the staying power of the system. Penalties in the game, as recommended by Lieberman (2010), are few and far between, and aim to incentivise good behaviour (playing the game), by removing the potential for greater rewards.

4.3.3 *Feedback*

Feedback refers to the players need for mobilising their own sense of worth and accomplishment (Abrams & Gerber 2013). More specifically, this Serious Game Mechanic aims to allow players to express their individual personality within the system

and be rewarded with a tangible feedback mechanism. Feedback mechanisms were identified as intervention functions in Section 4.2.2.1 along with being supported in literature as a successful technique for both empowerment and retention of knowledge, as detailed in Chapter 2.

Feedback as identified in the system is composed of two separate concepts. Firstly, the player must be guided on their achievement with respect to the learning outcome of the activity. This feedback will take the form of post-game text, illuminating how the player achieved during the session, and a hint as to how to solve the questions presented to the player, provided by the teacher. This feedback forms a core component of the study design and is illustrated further in the following sections.

In the context of game design, potential avenues for feedback: the player builds their character from scratch, when they first sign into the system. The player first begins their customisation by naming their character, and then choosing the type of animal. The theory behind building from scratch is that player becomes vested in the character, as it is an avatar reflection of their personality in the video game (Charsky 2010). As previously mentioned in Section 2.2.1.4, mechanisms whereby a player can exert a degree of permanent control and context within a game are important for a sense of accomplishment.

4.3.4 *Progression*

In the context of the game system, several mechanics are clear which may foster a sense of both development, and accomplishment; key components of progression. Each player of the video game intervention is completing an activity that intends to improve their reading comprehension skills. As such, it is important for the player to own a sense of accomplishment over time, with respect to their real world reading comprehension skill, along with their in-game progress (Lieberman 2010).

Each player is awarded a level as they progress through the video game. This level is a combination of game scores, along with reading comprehension scores. A player's level is also upgraded every time they interact with the game in a succession of days. Each time the player interacts with the video game, the player is awarded a game score. The score reflects how well the player performed the required actions during play

time. This score is presented to the player during and at the end of every game session. These scores are presented as graphs over time, such that the player can see the changes in scores as they progress through the game. As described in Section 4.3.2, tokens and points allude to the progression of a player over time, with respect to achievement. This progression forms part of the incentivisation of the player, to continue achieving (Abrams & Gerber 2013).

4.3.5 *Tutorial/Help*

As the player gains familiarity with the game mechanics, the relative difficulty of the game reduces, but still remains challenging to the player; this forms a core concept of video games, where experience with the game decreases the relative difficulty of play, disregarding mechanics that change in difficulty over time (Parkin 2017). Along with the relative difficulty of the game, the game content is adjusted by the teacher, on a day-to-day basis. Each player is monitored for their participation in this system, and the game content can be adjusted for each player by the teacher.

The player is also presented with their overall game achievement as their progress to the endgame condition: the planet Nautica. Each player is informed of how far they are from Nautica each day they play, and how long it will take them to get there. The player can interact with the system as long as they wish. However, as the system is composed of reading material, this material will be exhausted each day. The player must wait each day for new content for the game, with the intention of spreading out the achievement and effort of the player over a prolonged period. The forced waits also allow students to maintain a similar level of exposure to the intervention as their peers.

4.3.6 *Cut Scenes/Story*

The purpose of story elements of serious games is to integrate game mechanics and challenges into a broader context (Kiili 2005). More generally, the story of a game allows secondary meaning to permeate through player actions. The player has a second incentive to choose to play the game. While the purpose of this system is to encourage players to engage with SGMs, the role of storytelling is considered of some importance. A story mechanic was designed around the core game loop and is described in greater detail in Section 4.5.7.

4.3.7 *Tokens*

The system was designed with two forms of milestone unlocks. Firstly, players are assigned a level, which improves every time a player completes one-day worth of activities. This level appears next to their name and is accompanied by an animation and fanfare when their level increases at the end of a game session. The second form of milestone unlocks relate to the player unlocking new forms of hats for their character to wear during gameplay. While the hats provide no tangible benefit to the gameplay, the player is given a choice in how to decorate their character, and have a goal to unlock new hats. Every time the player reaches a new level, a new hat is randomly unlocked.

4.3.8 *Behavioural Momentum*

Behavioural Momentum concerns the speed and regularity of a person's free willingness to engage with a desired behaviour (Nevin & Grace 2000). This momentum can be therefore thought of, as a metric of how likely—conscious or subconsciously—a person will be desiring to engage in this behaviour. As this research relies heavily on behavioural analysis, the importance of the momentum of a person being inclined to engage in the desired behaviour is of importance. The context in which this momentum is achieved is tied to a variety of factors; most heavily in this system, are the use of tokens, levels, and rewards for the player, as previously described.

4.3.9 *LM-GM Summary*

As described in Section 4.2.2.1, the purpose of the BCW process is to develop a system through which behaviour can functionally be modified. It was identified that Education, Enablement, and Incentivisation were promising intervention functions. Game mechanics were explored using the GOP, and their applicability with identified BCTs was linked. Following this, the LM-GM model was used to bring together the three components of learning mechanics, game mechanics, and behaviour change techniques. With respect to BCTs, as previously described, Enablement can be facilitated through the LM-GM model of designing serious game mechanics; it is the purpose of Enablement, to generate a sense of agency for the player (Michie 2012), and as such, milestones and accomplishment mechanisms were detailed. As Education is in fact a

presentation problem, with respect to game design principles, the exact mechanisms through which reading comprehension skills are imparted is detailed in the following sections. Finally, Incentivisation was described as a core intervention function of the BCW process. Game mechanics were described with relation to Incentivisation, as potential mechanisms through which the player felt the need to continue with the activity, across the course the serious game.

Again, following the mapping of behaviour to game elements, the design of both a game and an administration portal was required. The following will now describe these two design processes.

4.4 Designing the Game

To test whether a serious game has an impact on the reading performance of students, either an existing game must be repurposed for use (serious or not serious), or a new serious game must be developed. As the extant literature provides few examples of video games for changing behaviours towards reading comprehension, the second option was selected, to fit the requirements of this study.

MiniNauts is a serious game package designed during this research project which aim to provide a structured and entertaining experience for students to practice their reading comprehension skills in a late primary school setting (Australian school grades 4-6). MiniNauts also aims to provide the teacher in charge of the class the facilities to operate an educational game in their classroom and be advised as to their class' performance.

MiniNauts is broken into two separate modules, whose purpose aligns to either the student or the teacher. The two modules are: administrative and diagnostic tool for teachers, and a video game system for students (MGS). Breaking the game into two separate components was a fundamental requirement for teachers, who felt that they needed time to observe student performance after the student had played the game, due to reasons of logistics and time management. The following sections will describe the development of both the diagnostic and administration too, as well as the serious game and its components.

4.5 *MiniNauts Video Game*

MiniNauts is a serious game which is administered through an online portal, accessible to teachers. Students interact with MiniNauts through their browser, and attempt reading comprehension activities through playful serious game experiences. The game was designed through close consultation with teachers where the game would be employed in their classroom. Thematic and genre choices accommodate as many of the student and teacher suggestions towards an idealised literacy game as possible, the genre and game type were chosen by the students themselves, during the survey.

MiniNauts takes the form of two mini-games, linked together by common reading comprehension content. The player plays through both mini-games in succession and is presented with feedback on their performance (dependent upon cohort) at the end of each mini-game. The purpose of the two mini-games is to present multi-modal learning techniques, with the purpose of allowing the player to interact with a piece of reading content, in two different contexts. The mini-games themselves are isometric 2D platformers, where the player is in control of an avatar, of which they can customise. Gameplay levels are randomised upon every playthrough. Reading comprehension elements are embedded within gameplay mechanics, and the presentation of the game, and reading elements are designed to closely interact with each other, without breaking immersion. The specifics of each of these game mechanics and components is outlined further in the following sections.

4.5.1 *Game Elements*

MiniNauts is currently composed of two mini-games that provide facilities for players to engage with reading comprehension activities. MiniNauts is broken into two separate mini-games. These games will be referred to as Game 1 and Game 2 from here onwards; MiniNauts does not present game mechanics between the mini-games.

4.5.1.1 Player Character Customisation

To encourage association between the player and the in-game avatar, a simple system of player customisation was devised. The work of Annetta (2010) points to game avatar customisation can increase game engagement. The feature of character

customisation also ties to the outcomes of the LM-GM analysis, presented in Section 4.3. Players are able to engage with two features typical of video games. The player is able to name their player character, and secondly, the player is able to choose the basic design of their character, along with several customisation options to their appearance as shown in Figure 4-5, thus following the LM/GM suggestion of Incentivisation and Enablement. Players are incentivised to customise their character.

Figure 4-5 Character Customisation

4.5.1.2 Mini-Game Features and Attributes

Each mini-game represents an atomic unit of learning and gameplay. This follows the recommendations of Koops & Hoevenaar (2012; 2008), which stress the importance of mini-games as didactic units. The mini-games will be arranged into a set, comprise of four discrete games. As the system of creating mini-games in this fashion has not yet been attempted, two games were chosen as a conservative number which allowed variety between the gaming experiences but did not overwhelm the designing of the entire system.

The set of mini-games are comprised of the following features:

- Each mini-game contains broadly the same thematic style.
- Each mini-game consists of similar mechanics to each other mini-game. This is to ensure the consistency of flow across the system.

- Each mini-game has different gameplay from each other mini-game. This is for two reasons. Firstly, to ensure that the games are interesting and varied enough to hold the player's attention. Secondly, this is to ensure that the system of mini-games is composed of distinct gaming experiences, and not levels of the same game.
- Each mini-game is composed of a different genre to the other mini-games. This is to allow different modes of expression in-game as to the literacy content, while also presenting variety in the system for the end users.

The mini-games can be considered as separate entities from each other, but form part of a system of teaching. Each mini-game is playable up to a set length of time. Retaining the premise of a conceptual mini-game, as described by Koops & Hoevenaar (2012) is important; the game will end after a certain period of time, but will be scalable in game difficulty over the period of play. Figure 4-6 demonstrates the way in which activities are presented to the player. This is to ensure that players are not presented with challenges that are well beneath their skill level, to aid motivation in using the system. Furthermore, each game is constructed in relation to Kong et al. (2010) principles of good game design. These principles are designed to encourage interaction, agency and motivation to use for an educational system; these are identified as requisite components of learning (Kong et al. 2010).

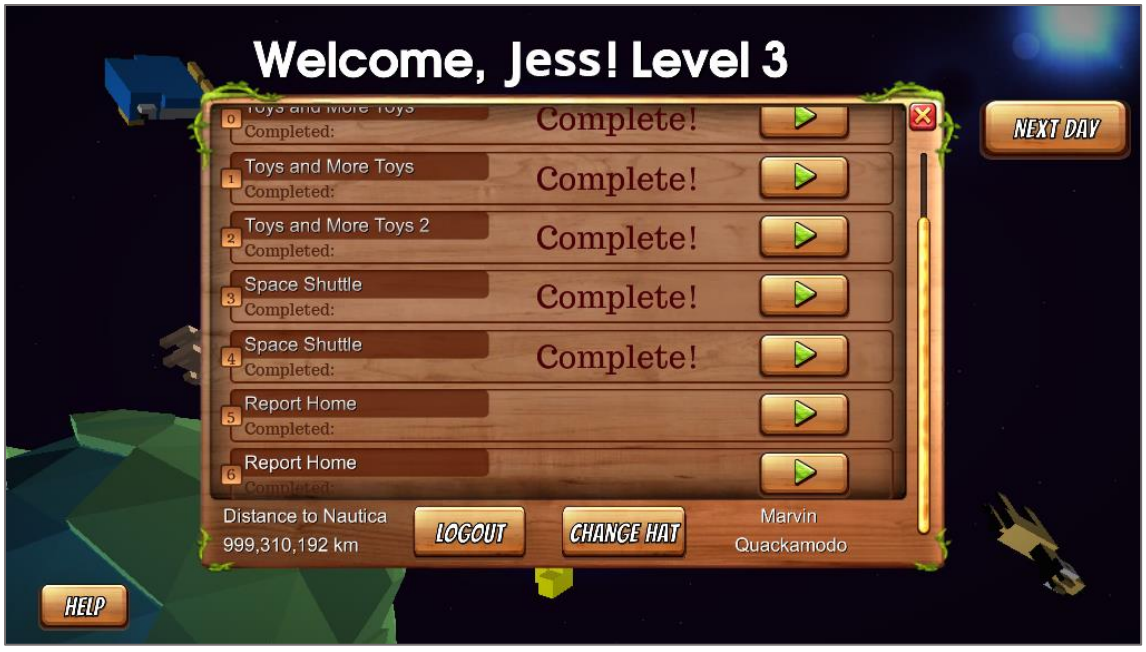


Figure 4-6 Activity Selection Screen

The MGS also presents the concept of three resources. These resources, named readium, thinkium, and guessium serve three purposes. Firstly, they act as reward tokens for the player once they complete a reading comprehension activity, this relates to the GM of Tokens, as described in Section 4.3.7. Secondly, these resources are markers for the system to record player achievement over time; the GM of Progress as described in Section 4.3.4. All students gain the three resources as they play and mark the passage of progress as they play with the MGS. Finally, the three resources can be spent as tokens. As mentioned in Section 4.3.7, unlocks and milestones were identified as possible game mechanics. The player may choose to spend their tokens in the game store, expanded in Section 4.5.8.

The MGS system attempts to enable players to engage in reading comprehension activities, along with incentivising them to interact with the system for as long as possible, as described previously in Section 4.3.

4.5.2 *Multiple-Choice Quizzes*

MiniNauts relies heavily on a multiple-choice quiz structure, to present learning content to the user. MCQs have numerous benefits for both educational providers and students. Firstly, MCQs form a core component of the Australian curriculum and current testing practices for reading comprehension in Australian schools. This level of familiarity works to the advantage of both educational providers, along with students. Secondly, MCQs are a common game design element with respect to serious games, and have been presented in numerous fashions (Gee 2005). These facts point to MCQs providing a compelling way to package educational content into a video game context; these become context clues for the player, to aid in the comprehension of text The manner in which these MCQs are presented to the player, are now of importance. Prensky (2005) describes the challenges inherent in digesting content into units deemed too small, as such, MCQs are perhaps the smallest unit of educational content and testing practicable. However, the following section describes a unique method whereby MCQs are repurposed into game elements, and remain tightly coupled with gameplay.

4.5.2.1 Reading Comprehension Quizzes

Multiple-choice quizzes, with respect to reading comprehension take a basic form. Firstly, a portion of text, whether a paragraph, essay, or single sentence is presented to the reader. It is the task of the reader to engage with the text presented, in its whole form. Following this, a question is posed to the reader, with the form of the question posed to be answered by two, to multiple answers. The difficulty in selecting the correct answer is dependent on both the reader's ability to reason an answer, along with the dissimilarity of the answers from one another (at the discretion of the teacher). If the answers are dissimilar to one another, the chance of choosing a correct answer increases, as long as the reader has the capability to identify the most probable answer either deductively from the text, or inductively from the answers provided (Rupp et al. 2006). As detailed in the following sections, the basic structure of the MCQ is used in both *Game 1* and *Game 2.*

4.5.2.2 Game 1 Word Substitution

The first mini-game presented in MiniNauts requires the player to guess the missing word in a sentence. This technique of word substitution is a well-founded component of reading comprehension testing, and training (Keenan et al. 2008). The purpose of this technique is to determine if a reader can make meaning from a sentence, even if the sentence is incomplete; this demonstrates whether a reader can use multiple sources in a sentence to make meaning. The process for generating word substitution activities is completely automated. Since there is a correct and incorrect answer for word substitution activities, the lack of ambiguity leads to a process whereby a system can dynamically remove words from sentence and present players with possible answers. It should be noted that whilst words form such as verbs and nouns cannot easily be substituted for one another, adjectives and adverbs can present the possibility for confusion. For this study, the system automatically substituted words it identified only as nouns and verbs. Possible word substitutions were generated from an online list of synonyms, and antonyms, with a varying degree of randomness for the word list. This randomness was shared amongst all participants, so there were no words which some students received that others did not on a per-activity basis. Table 20 presents possible

word substitution options that the MGS would generate from a source sentence; the first for a verb, and the second for a noun.

Sentence	Missing Word Sentence	Possible Words
The alien spaceship flew silently through the air, landing in the grassy field with barely a thud.	The alien spaceship ___ silently through the air, landing in the grassy field with barely a thud.	Flew – Jumped – Rode – The – Landed
	The alien spaceship flew silently through the air, landing in the grassy ___ with barely a thud.	Floor – Field – House – Trampoline – Sky

Table 20 Word Substitution

4.5.2.3 Game 2 Statement Validity

The second mini-game encourages the player to use critical reasoning to determine whether a textual statement is correct with respect to a body of text. The following quiz presented in Table 21 details the basic reading unit of *Game 2.*

Text	Statement
Harry poured milk onto his cereal. He really enjoyed breakfast, as it was the calmest part of his day. As he ate a spoonful, he contemplated what he would get up to this coming weekend.	A) Harry was in a grumpy mood
	B) In the story, it is the weekend
	C) Harry has a busy day ahead of him
	D) Harry was eating cereal for breakfast.

Table 21 Sample Reading Comprehension Quiz

It is the reader's task to determine the validity of each statement, through abductive reasoning and whittle down the choices until one correct choice remains (Rupp et al. 2006). This process of elimination through multiple options is a form of context clue, and provides a contextual point whereby meaning can be made for the learner with respect to the context of a statement against some supporting text. In this

example, both Options C and D can be considered as valid statements (with Option D the 'truest'). A reader would read the paragraph text, and then choose from the Options which seems most likely. What is absent from most paper based MCQs is the immediate feedback on whether the reader's chose correctly or incorrectly and why. Electronic tests have the capability to perform immediate feedback but remain as unappealing as paper-based quizzes. Game 2 is an attempt to alleviate that problem and provide a space for players to engage with reading comprehension through assessment, and immediate feedback.

4.5.2.4 Game 2 Point of Inference

Complementing the ability to use an MCQ as check for statement validity is the ability to draw connections between paragraph text and factual statements. This process can be referred to as inference, and is a critical component of reading comprehension (Anderson & Pearson 1984) The purpose of this activity is to locate the place in-text where a true statement's meaning can be derived from logically. The text presented in Table 22 outlines the process through which the point of inference can be gleaned.

Sentence	**Statement**
Harry was a happy child.	Harry is an animal lover.
One of his favourite activities was walking in the park with his dog Buster.	
Harry mistrusted his friend Sally, since she always lost Buster when she held his leash.	

Table 22 Inference from statements

The reader's task is to determine through gameplay, which sentence provides meaning and context for a given statement. In the case of the preceding example, this task is made easier by the fact that the sentences are demarcated into separate sections of meaning, with the subject of the sentence remaining consistent through each sentence (Harry). In the example presented in Table 22, the statement *"Harry is an animal lover"* is deemed to be true. Therefore, the reader looking through Option A concludes, that the sentence does not confirm this statement to be true, as the subject is not the same. Option C mentions *Harry's* friend *Sally* but makes only a small reference to *Buster* being lost,

and how this makes *Harry* mistrust *Sally.* The reader could infer that the fact that *Harry* does not like it when *Buster* is lost, that this implies he is an animal lover by implication. However, Option B presents the most compelling case that *Harry* is an animal lover, stating that he enjoys taking *Buster* for walks in the park. While neither Option B nor Option C explicitly maintain that *Harry* loves animals, the reader can reason which option to be the 'truest' in this circumstance. In the wider context of reading comprehension as a tool for literacy, the ability to identify the place in text where meaning can be derived from forms a core component for functional reading comprehension (Yuill & Oakhill 1988). Meaning making through the use of clues is also a form of context clues, whereby the presence of information relevant to another piece of information can help someone make meaning of a piece of text (Rupp et al. 2006). Game 2 makes use of points of inference activities to provide a gamified learning space where players can practice inference skills. This process is described in greater detail in Section 4.5.5.

4.5.2.5 Scoring

Scoring plays a vital role in the MGS. Keeping track of the player's individual performance during play, and presenting it back to the player, while also recording it for analysis is crucial to this study. Therefore, suitable metrics for determining player scores, and how they related to the various components of the MGS was a key consideration.

Scoring in MiniNauts are split into two categories: reading performance, and game performance. As previously described, reading performance can be measured in certain terms with respect to reading comprehension, if the activities presented to the player follow the structure of an MCQ. Game performance in MiniNauts is also measured in a similar binary manner, where individual actions in the game, incur a positive or negative penalty to the game score, with the final score reflecting the overall performance of the player over the course of the game. The player's score is presented at the end of the level, along with statistics of how they performed, and which actions they correctly or incorrectly chose. Game 1 is intended as a classroom activity that enables students to infer meaning from a piece of text in a structured and systematic manner. The game itself is a repurposing of the multiple-choice quiz format that is

popular in many Australian schools, particularly for NAPLAN testing. As has been previously stated, assessment activities can demonstrate not just the state of a player's knowledge or reasoning, but also act as an activity of learning and practice itself. (Stiggins 2002).

4.5.3 *Game 1*

The first game presented to the player when starting a new MiniNauts session, is that of Game 1. The purpose of Game 1 is to allow the player to connect missing words to incomplete sentences. This process of word substitution is detailed further in Section 4.5.2.2. Players engage with Game 1 via an isometric 2D perspective and play the game primarily with a mouse. Players may play Game 1 for as long as they like and are not penalised for taking longer to play the game. Correctness of actions in the game is the primary marker for score, which is itself a reflection of success at the game. The defining mechanics of the game are the juggling of player resources to perform the character word substitution, while also avoiding incoming enemy attacks that may hinder the player's progress. It is important to note that although enemies may attack and potentially destroy the player ship, the player is not penalised for these actions, and only their performance in the reading activity is the measure of success. This is mentioned to the player in the opening introduction to the game.

Figure 4-7 *Game 1* gameplay.

4.5.3.1 Gameplay

The player is in control of a small spaceship, that flies into the play zone at the beginning of the level. The player is equipped with a tractor beam, which allows the player to grab any object in the space close to the player ship. Across the level, are several scattered crates, with a word printed on the top. The player may choose to click on a floating crate, and this crate gets added to their ammunition store, represented as a trail of words behind the ship, as seen in Figure 4-7. This ammunition store follows the player around the screen, and it is clearly visible to the player which word they currently have selected and can shoot. As can be seen in Figure 4-7, the sentence reads: *Cassie rolled over in her bed as she felt the sunlight _?_ her face.* The player must choose which word suits the missing word in the sentence best from their ammunition store and shoot this word back at the enemy ship. If the word was correctly shot at an enemy ship, the enemy will be destroyed, and drop 1 unit of guessium. Across the level, there is guaranteed to always be a crate with the correct word to solve the challenge, and it is the player's job to shoot this crate at an enemy, to fill in the missing sentence. As previously mentioned, the player oversees juggling the word resources, and an incorrect firing of the word will result in the word fired bounce off the enemy ship. While the player is not penalised for shooting an incorrect word at the ship in terms of gameplay, their score is lowered. Players may also pick up previously destroyed enemies and asteroids and hurl them at the enemy, to delay an incoming shot to the player. The player may be destroyed by incoming enemy missiles, but they have the option of replaying the activity straight away.

Circling and moving around the player are enemy ships. The ships are constantly shooting at the player and attempting to evade the player's word missiles. Word missiles are words shot out of their ship, and if they hit the enemy, the enemy is either destroyed if the word is correct, the enemy ship is destroyed. The player must complete at least one of these activities to complete the level, dependent upon the teacher's choice in activity design during the content setup phase prior to players playing.

4.5.3.2 Scoring

Players are not informed as to the meaning of the text as they play. This is to ensure that while reading, they are successfully making logical guesses as to what word could be placed in the sentence. This method of guessing is intended to challenge the player's reason abilities, and as such, they are awarded the resource of guessium (described in Section 4.5.1). This resource is used to count their final score toward the activity and is shown to them at the end of the activity session, in the final score screen.

As previously mentioned, the purpose of MiniNauts is not to punish incorrect actions, but to reward correct ones. It is with this in mind, that a scoring system which attempts to capture all relevant actions the player took in the game, and then quantify them into a recordable metric was developed. These metrics are referred to in Table 23.

Scoring Item	**Description**
Guessium Collected	Each level provides a deterministic number of guessium that a player can collect. Each crystal of guessium is dropped once the player successfully completes a sentence.
Combo	The number of enemy PenGuy ships destroyed without taking enemy damage, represented as a multiplication of score.
Time spent playing	The time spent playing the activity represents how long the player spent engaging with the activity. This metric alone cannot measure individual performance, as players may struggle with game mechanics and not reading mechanics.

Table 23 Game 1 Score Metrics

As described in Table 23, scores reflect achievement, and as such, there are no negative modifiers to individual player score. Players simply gain a higher, or lower number of points at the end of the level. Players may choose to repeat the level for a higher score, and as such, be able to purchase cosmetic items for their character.

4.5.4 *Interlude*

Following the successful conclusion of Game 1 in the activity set, player score is presented to the player on a summary screen. This screen visually represents how the student scored during the game, but dependent upon Test Groups (refer to Section 5.3), may display more information for the player, such as feedback, detailed in Section 4.5.6, and displayed in Figure 4-8. All players are presented with information as to whether they levelled-up or not, and how many resources they collected during gameplay.

A level-up is defined as gaining the maximum total of guessium from the enemy, while not suffering from a single hit from an enemy missile. A level up can potentially occur at the end of every mini-game, and as such, a perfect playthrough of both mini-games can present the opportunity for two level-ups per game.

A short cut scene is then presented to the player, transitioning their spaceship from space, down to a planet; it is at this point that Game 2 commences.

Figure 4-8 Game Interlude

4.5.5 *Game 2*

The second game of the MGS presents players with the opportunity to infer meaning from a given piece of text in a gameful manner. The following sections will outline the specific gameplay details and design choices that enable this.

Game 2 presents a different game style when compared to Game 1, that is, a two-dimensional action game, which relies on character movement and combat with interfering enemies. As mentioned in Section 2.2.2.6, it was identified that thematically linked, but stylistically dissimilar mini-games could be linked for a serious game purpose. It is in this case, that Game 2 is linked to Game 1: in this manner, the player has contextualised the purpose of Game 2, with respect to Game 1 through a transition cutscene of the player's ship landing on the planet, along with being presented reading content that was related to Game 1. At this point, the player engages in a different manner of play, but is still working in the epistemic frame of the same reading content from Game 1. Recall that reading content is shared between the mini-games on an activity basis, and a hypothetical Game 3 could be inserted into the activity to expand the gameplay further, but still work with the same educational content of Games 1 and 2.

4.5.5.1 Gameplay

The purpose of Game 2 is to present players with a playful semi-sandbox environment, where they can explore the landscape, while completing activities. The player moves along a 3D landscape and controls their MiniNaut. The MiniNaut is stranded on an enemy planet and must refuel his/her ship so that they can continue their journey to Nautica. The player must refuel their ship by destroying the correct form of crystals and carrying their pieces to their stranded ship which needs to be refuelled.

Reading content is presented to the player as a popup screen that is open and the start of the game and can be reopened at any time and viewed by the player. The reading content is a paragraph of text, which the teacher has input into the MGS, during the Content Setup phase of administration, as detailed in Section 4.7.4.1. This paragraph text is intended to give context to player actions during gameplay. The player's score is a combination of factors described in the next section.

The gameplay is split into two separate phases, the first is the reading phase, where players collect readium, and the second is the thinkium phase, where players collect thinkium to refuel their ship. Both sets of crystals are needed to complete the activity and leave the planet. The gathering of resources is once again the primary focus

of the mini-game, this mechanic is shared between Game 1 and Game 2 for the purpose of continuity.

Figure 4-9 Game 2 Gameplay

In the first gameplay phase, players choose which crystal is the correct fuel type for the downed player ship via the association of a textual description above the player's ship and the crystal (Figure 4-10), demonstrated in Figure 4-13. Players must collect pieces (Figure 4-11) from the readium or thinkium crystals and return them to their ship (Figure 4-12); players choose the correct crystal to destroy by reading the popup text on the bottom right of their screen.

Figure 4-10 Readium Crystal

Figure 4-11 Thinkium Crystal Pieces

Figure 4-12 Returning thinkium Crystal Pieces to the MiniArk (player ship)

This textual association is detailed further in Section 4.5.2.3. In the top left corner of Figure 4-9, a question is presented to the user in yellow text. These questions relate to possible answer statements, present in readium crystals, shown in Figure 4-10. Players must destroy these crystals and return their pieces to the player ship. Identifying the correct crystal to destroy is the targeted activity of this phase, and involves the player strategically performing attacks on the crystal, while dodging incoming enemy penguin attacks. The player can receive three hits by a penguin, before they are killed in the entire game. Player health information is provided via the on-screen heads-up-display. Players may also pick up items that are dropped by enemies if they are killed by the player. These include armour, which increases the player's health by one per piece, and, speed boost items, that make the player move around the map more quickly.

Figure 4-13 Thinkium Crystal Text

The second phase of Game 2 is for players to collect thinkium crystals in the game world. Players can collect either thinkium or readium crystals in any order that they choose. Thinkium crystals are collected by reading popup text when hovering the mouse over relevant crystals (Figure 4-13); whose attached sentence is the sentence from which readium crystals were linked to in text. Recall points of inference, as detailed in Section 4.5.2.4 where a reader must decide where meaning is derived; this inference linking process is intended to complete the cycle of comprehension, by linking back the correct answer to the correct sentence in which the correct answer was derived. As in phase 1, these associations are generated by the teacher during the content setup phase. The player is awarded readium, and thinkium for their efforts.

Enemy penguins attack the player and attempt to steal crystals from them, halting the speed of the player's refuelling efforts. It is the purpose of these enemies to carry off player's crystals and dump them in the surrounding ocean. Part of the game appeal of Game 2 is juggling reading, combat, and transporting crystals at the same time. Players are free to pause the game as they choose, and to evaluate which answers to pick, so as to not be overwhelmed. The player may interact with the game world in two ways:

- Quacking, to destroy a PenGuy or a crystal, three successful quacks are needed. Figure 4-15 shows the player in the middle of combat

- Grabbing game world objects. The player may grab any close available game object and throw it anywhere. If a player chooses to throw an object at an enemy, the enemy may take damage and drop coins; shown in Figure 4-14. In this manner too, players also can transport their crystals to the spaceship. Players cannot attack enemy characters unless they have dropped any crystals that they are holding. In this way, combat and progressing the refuelling are linked. Players can also choose to pick up and rescue imprisoned MiniNauts, and send them back to their space ship, for extra points. These imprisoned MiniNauts are randomly scattered around the map and serve as a tertiary goal for the player.

Figure 4-14 Game 2 Throwing Objects

The question and answers used in Game 2 are derived from each MCQ the teacher generated in the Content Setup phase. Once the player successfully refuels their ship, the player can choose to enter their ship, and fly off-world.

Figure 4-15 Game 2 Combat

4.5.5.2 Scoring

Game 2 implements a similar scoring mechanism as to Game 1. Each level aims for the player to score as highly as possible, with score reflecting the gameplay and reading comprehension aspects of the gameplay session, these metrics are referred to in Table 24.

Scoring Item	Description
Coins Collected	Each level provides a deterministic number of coins that a player can collect. Each coin is either randomly located in the world or pops out of a PenGuys when quacked at.
PenGuys Destroyed	The number of PenGuys who are destroyed counts towards the players score.
Time spent refuelling	The time spent refuelling the ship reflects how long they could keep the PenGuys at bay. A lower time results in a higher score.
MiniNauts saved	The player must pick up frozen MiniNauts and return them to the MiniArk, providing points for each MiniNaut saved.

Table 24 Game 2 Score Metrics

As is the case for Game 1, the player's score is a direct reflection of their ability to perform reading comprehension activities, but also includes tertiary goals. If the player's score is high, it means that they have performed the reading comprehension activities successfully. Thinkium and readium crystals act as score multipliers, such that even if the player completed as many of the tertiary tasks as possible, they would not score as highly as a player who completed only the reading tasks. The player with the highest score however is the person who completes all primary and tertiary activities.

4.5.6 *Feedback*

As described in Section 5.3, the feedback component of the MGS is only provided to students in TG3 (refer to Section 5.3). This is to ensure the validity of the test condition. For the remaining two Test Groups, no feedback will be presented to the player following a play session. The game will explain performance through score and success criteria, as in regular video games and serious games.

Between each mini-game, the player is presented with exact feedback on how they performed. This is an attempt to provide students with formative feedback on their play experience, a known way knowledge can be formalised in memory (refer to Section 2.2.1.3). The feedback takes the form of statistics on how well they performed, including their score and high-score, along with their play-time. This form of feedback is well supported by literature (Abrams & Gerber 2013; Zebel et al. 2013; Hattie & Timperley 2007). For example, in Game 2, the feedback for a player could say:

- You did really well!
- Remember to pay attention to people/places/things to find out what's going on.

- He enjoyed playing in the park with his dog **Spot**, and his friend **Sally**.

Figure 4-16 Game 2 Feedback

The game content is marked with points that show meaning in the text, in the case of Figure 4-16 the proper nouns. These relationships are established in the Teacher Portal content management system. These hints are written by the teacher at the time of establishing content, as mentioned in Section 4.7.4.1. The purpose of both the quantitative presentation of how well the player performed, along with hints provided by the teacher with respect to the exact nature of the task aim to provide participants of TG3 with both elaborative and formative feedback at regularly scheduled intervals, as supported by (Abrams & Gerber 2013; Hattie & Timperley 2007). Whilst MiniNauts aims to provide as many feedback opportunities to players as possible in TG3, it must be noted that all feedback is automated. Teachers may ahead of time include hints for the player, but even further customised feedback is not present, as this would lose the autonomous characteristic of content presentation; a form of summative assessment (Chen & Michael 2005; Sarroub et al. 1998). In this sense, a balance between quantitative, and elaborative feedback, against the backdrop of a fully autonomous serious game system is struck. As shown in Figure 4-16, scores relating to player actions in MiniNauts

are displayed to participants through an interactive dialog at the end of each game session.

4.5.7 *Narrative*

A key element for successful serious games is the use of narrative as a motivational and attention-grabbing device (Amory & Naicker 1999). As such, MiniNauts contains a narrative that progresses over the course of the intervention period. MiniNauts are a band of wandering animals searching for their lost home planet of Nautica. The MiniNauts live in a generation ship called the MiniArk, which has been travelling for many years around the galaxy, searching for the lost planet. The player assumes the role of one of the MiniNauts (a duck), called Marvin who is responsible for finding lost MiniNauts and bringing them back to the MiniArk. The story of the game is progressed through short dialogue cut scenes, presented at the start and end of each level. The player is alerted to their game progress via the progress screen presented in Figure 4-17, along with game cut scenes between the player character and NPCs.

Figure 4-17 Progress Screen

To contextual the individual atomic gameplay experiences of the player into a wider framework, it was deemed necessary to create a wider goal for the player to aim for; reaching Nautica. While this goal is not the purpose of the game it serves to frame player actions within the game, and provide greater meaning to their actions.

4.5.8 *Game Store*

To spend the previously mentioned resource tokens of readium, thinkium, and guessium, a game store is present in the main menu of the MGS. The store allows the player to spend a certain number of their resources for new clothing items to update their character, along with the ability to add epithets to their name. The purpose of the store is to allow players to perform meta-game actions that extend the use of the game beyond each play session. (Kluger & DeNisi 1996). The purchasing of cosmetic items is considered optional but is presented to the player as a goal that they could work towards, if they wish to spend some of the points they earned during a gameplay session.

4.5.9 *Help*

Players of MiniNauts are presented with several help screens, scattered throughout the system. Players have the opportunity to click a help button, and pause the game if necessary, to review the goal of the game, and any controls required. The help system is intended to give players an opportunity to solve problems without asking for their teacher's attention in the first instance; this mechanic was elucidated from the LM-GM analysis phase presented in Section 4.3.5.

Figure 4-18 Game 1 Help Screen

4.6 *Control Group Games*

As the purpose of this study is to determine the efficacy of providing students with feedback at regular intervals and determining how successful a gamified system for practicing reading comprehension, is a condition in which no gamified elements are presented is necessary to serve as the control in the experiment (see Section 5.3). For both Game 1 and Game 2, gamified mechanics revolve around interacting with literacy objects as mechanics. The control games will provide the intervention to students in the following manner.

As previously described, ethical considerations are imperative for school-age children, especially those of equity. It is imperative that each student should receive an appropriate opportunity for engaging with educational content in the classroom. With these considerations in mind, the following modifications and substitutions to Game 1 and Game 2 were designed to ensure an equitable experience is provided to students, while also exposing them to different modes of content delivery.

4.6.1 *Game Modifications*

Each player, at the beginning of a play session is presented with an MCQ, if they are placed into the control condition. The MCQ is the same in content, as the gamified reading elements of the previously described games but are presented in a tradition question answer format. Once the player completes the MCQ, they are placed into a modified version of the game. A sample of how MCQs are displayed to the player are shown in Figure 4-19.

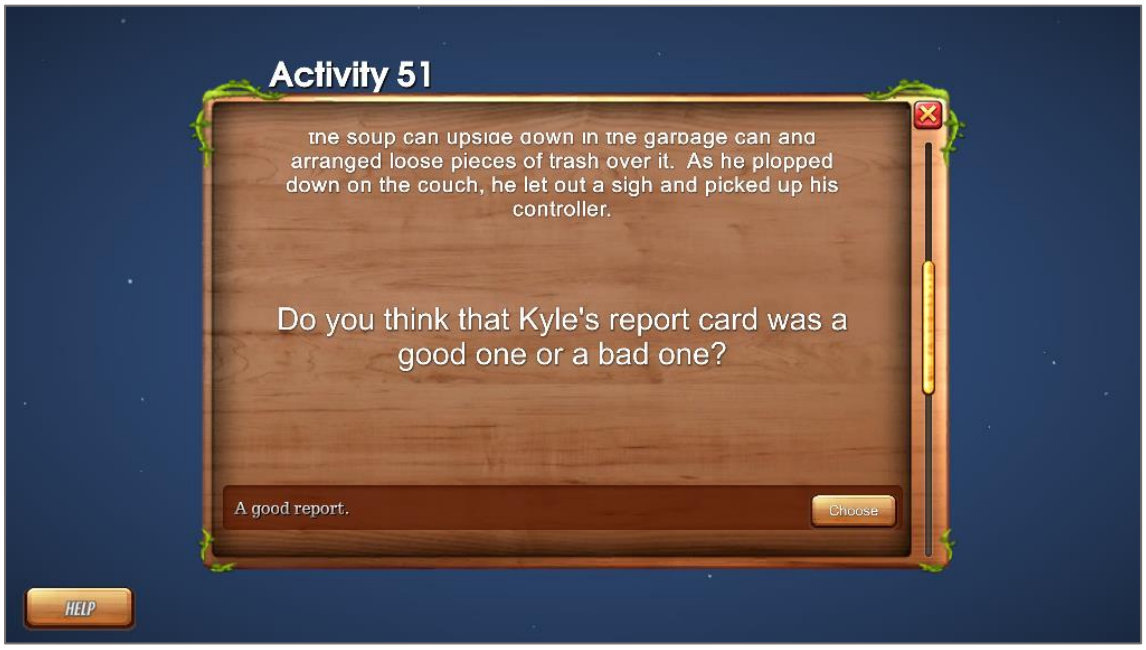


Figure 4-19 Control Game MCQ

4.6.1.1 Game 1 Modifications

The player is presented with simple maths problems to solve rather than performing sentence completion. Each math problem is either an addition, subtraction or multiplication problem that is randomised. The player progresses through the level in the same manner in which their peers in TG1 and TG2 progress, with the capacity to earn the same scores as the other players. They are also rewarded with guessium, for each correct answer. Maths problems were chosen as a substitute for reading comprehension questions for two reasons. Firstly, to maintain consistency in game mechanics, question/response activities were needed to minimise the differences between the control group, and the intervention groups. Maths questions were chosen as an educational substitute to reading comprehension and suggested by the teacher participants as a way for students to engage in educational content. Secondly, ethical research standards in Australia regarding students states students must not be disadvantaged when compared to their peers if an intervention model is used. The presentation of a second educational factor in the control group added to the weight that the research was not disadvantaging students unfairly. The choice of mathematics may be a confounding factor on motivational metrics for the control group, and this will be assessed in Chapter 6.

4.6.1.2 Game 2 Modifications

For the second game, the player must solve spatial puzzles to complete the refuelling process. The gameplay remains the same between the control and two game conditions, but rather than linking statements to questions, the player must solve several versions of spatial and pattern puzzles. These puzzles typically take the form of complete the missing sequence questions, where fuel pumps have their hovering sentence text replaced with symbols. The player also can no longer can pause the game and view the current game text and game questions, as these are removed from the gameplay and serve no purpose.

4.7 System Architecture

The MGS described throughout this study relies on the automatic collection of participant data via the internet. The following sections will describe in detail the process of serializing and storing any data captured, along with the format with which the data is stored and transmitted.

4.7.1 *Distributed Application Structure*

MiniNauts uses a client-server model as the basis for its system architecture. Each game instance draws upon resources from a central server, located on the NECTAR Research Cloud (Nectar Directorate 2017). The game requests player information, their statistics and progress from the central server and is delivered to the game client, via a HTTP accessible web API. Once the player has performed a permanent action, for example, registered a new account, or finished a game session, the game client performs

a HTTP request to the server API, sending data back to be processed and stored. This process is described visually in Figure 4-20.

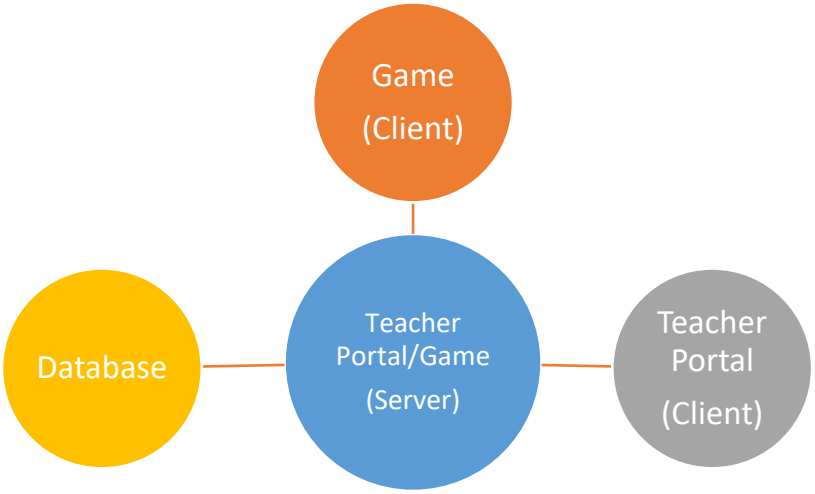


Figure 4-20 Client/Server Model

4.7.2 *Game Client*

Each game client runs in a web browser, as a WebGL HTML5 application. These applications run natively in modern web browsers, and do not require anything beyond a persistent internet connection to be played. The start-up execution procedure for the MGS is described in Figure 4-21.

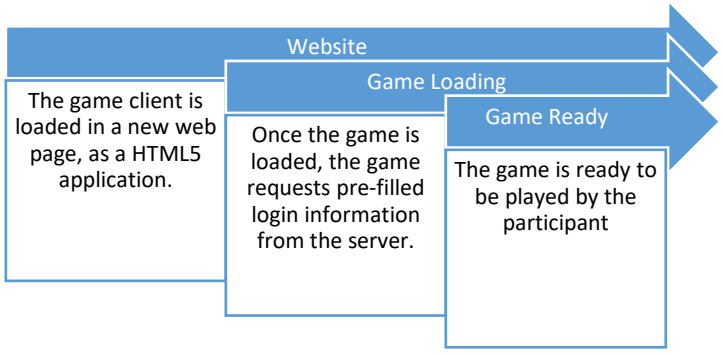


Figure 4-21 Game Start-up Procedure

Once all required components have been downloaded to the player's browser, the player can begin to interact with the MGS. The game client is approximately 15 megabytes in size and requires minimal bandwidth to maintain the ongoing connection past the initial game load.

4.7.3 *Server*

The central server used in this study resides on a managed virtual machine, located on the NECTAR Research Cloud (Nectar Directorate 2017). The server runs the Ubuntu 16.04 operating system and is headless for the majority of the research study. The server runs an Apache 2 web server, which provides HTTP and SFTP access from internal and external connections.

4.7.3.1 Internal Access

Read/write access to the server is protected via SHA-256 encryption, which is monitored via the researchers of this study. All files and data are accessed to and from the server via SSH when requested over the World Wide Web.

4.7.3.2 External Access

The server is publicly accessible to the World Wide Web as a regular web server and web service. All web requests are made through the server to PHP files, which handle data throughput. The server does not expose the database used in this study directly. The server uses SSL encryption for any client/server requests which contain any user information. These requests apply to both the Teacher Portal and MGS.

4.7.3.3 Database

The database used in this study is a relational *MySQL* database hosted on the previously mentioned NECTAR virtual machine. All sensitive data is salted and hashed in the database.

4.7.4 *Teacher Portal*

The Teacher Portal (Figure 4-22) allows any teacher with access to the system to administer MiniNauts activities for students and gain insight into how each student has performed. The Teacher Portal is composed of compartmentalised components; intended to provide different functional for the teacher.

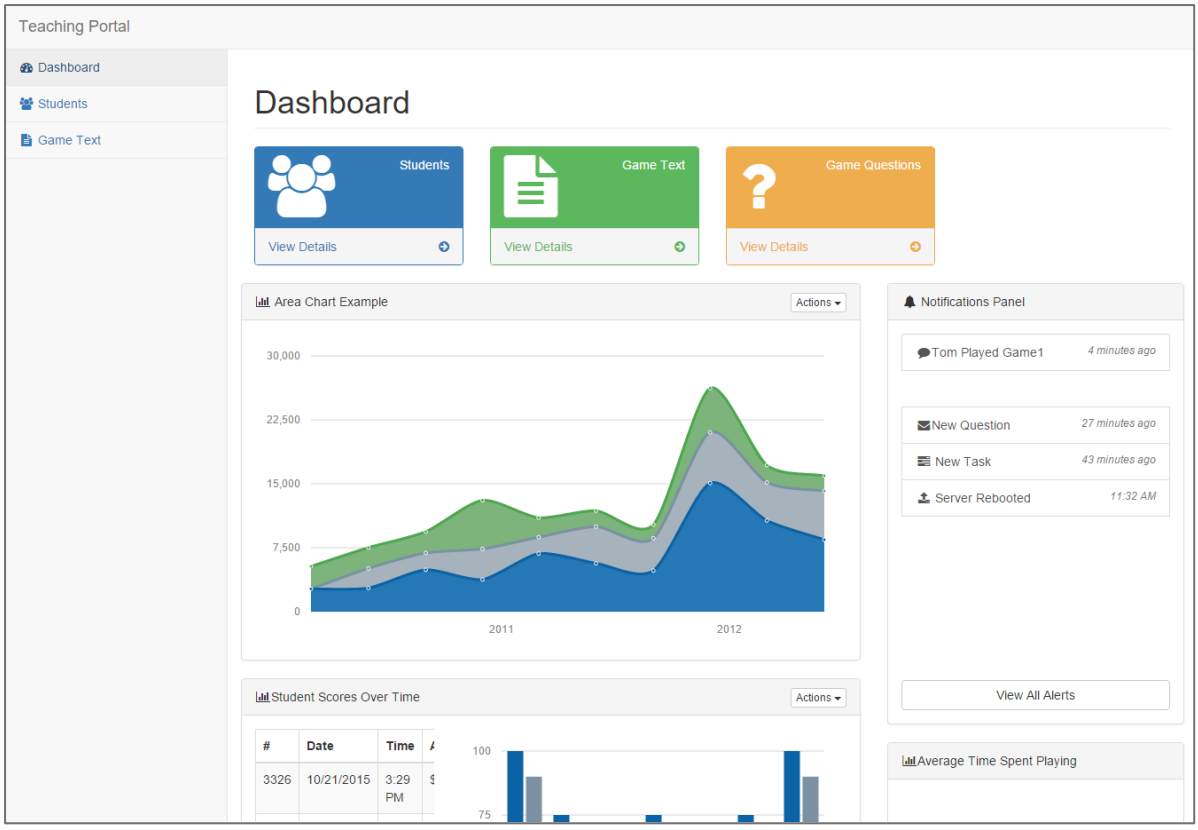


Figure 4-22 Teacher Portal Administration

4.7.4.1 Content Setup

A core principal of MiniNauts, is that of content-agnosticism. The purpose of content-agnosticism is twofold. Firstly, content for educational games can be strongly or weakly coupled with game mechanics. For practical purposes, particularly with modern curriculums, one size may not fit all. Therefore, MiniNauts aims to remove the content of the reading comprehension activities from the actual mechanics of the game. This means that the educational content is decoupled from the gameplay of the video game. As previously discussed in Section 2.2.2.6, a strong coupling of content to the video game may increase enjoyment, but limits replayability and increases development burden.

With the requirements of the software in mind, it is the purpose of the Content Setup portion of the Teacher Portal to enable teachers to perform the administrative task of designing reading comprehension activities, for use in MiniNauts. Figure 4-23 illustrates the layout of the Content Setup main page of the Teacher Portal. The page is broken into two sections; the left, is composed of paragraphs of text that are to be read

by the students. To the right, are MCQs which are presented to the students during gameplay, dependent upon control group.

The content setup is accomplished with the following procedures:

1. The teacher enters the core text for the video game into each of the four paragraphs input fields. The website automatically breaks down the text into individual sentences that can be manipulated. A reading age score is calculated by the system, using the Hertfordshire Reading Test methodology; the use of this score was a minor requirement requested by teacher participants.
2. The teacher enters one to several statements for each paragraph, in the corresponding table to the right of the paragraph text. A statement is a sentence which relates to the paragraph text in some way, much like the individual options of a multiple-choice quiz. For example, if the paragraph text read:

An example of some MCQ content is presented in Table 25.

Paragraph Text	**Statements**
"John was a naughty child. He enjoyed breaking all of his toys."	John was a good child.
	John had a lot of friends as a child.
	John enjoyed being destructive.

Table 25 MCQ Example Content

Each statement can be regarded as either true or false in relation to the context of the paragraph story. Each statement can also be verified for its truth or falsehood by a corresponding sentence in the text.

3. The teacher establishes these relationships with a simple click and drag mechanism, and the video games are automatically updated to reflect the new content and relationships.

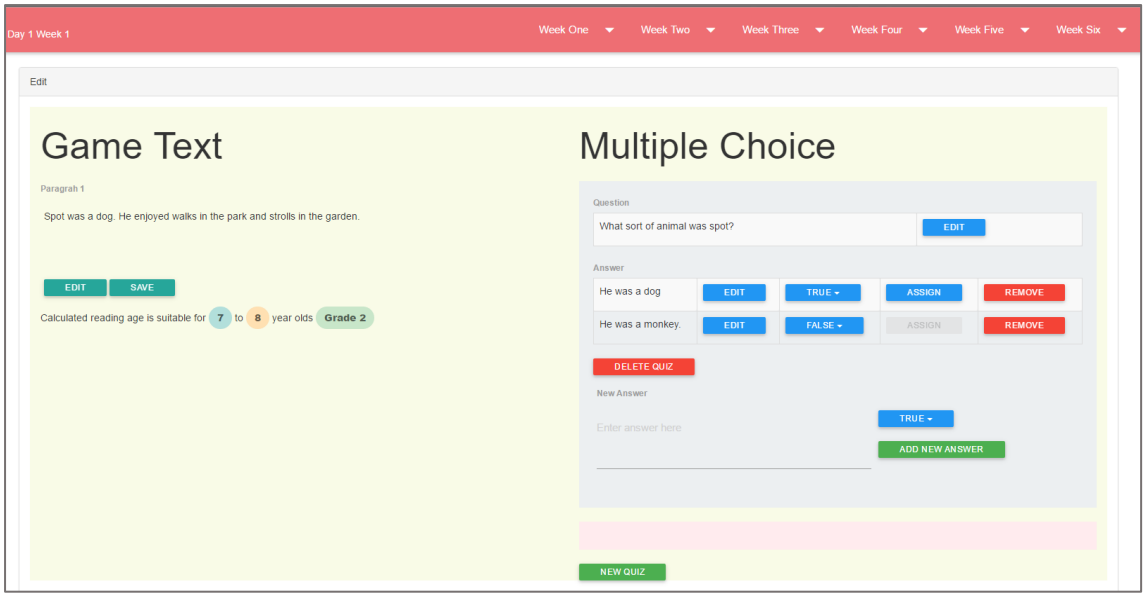


Figure 4-23 Teacher Portal Content Setup

Recording student performance in educational games is an important way of gauging the efficacy of the video game as a learning aid. MiniNauts automatically collects various data points on the play of each student. Statistics such as play time, correct and incorrect actions as well as high scores are all recorded. It is hypothesized that as the primary goals of each video game is to succeed at a reading comprehension task, success at the video game may correlate to an increase in reading comprehension.

Along with individual student performance, a teacher is responsible for the class' overall reading performance. It is with this in mind, that the Teacher Portal contains the facility for teachers to view the overall trend of the class when it comes to the metrics of the game (refer to Figure 4-22). Change over time, along with average score and time are displayed to the teacher, as a tool for gauging the overall performance of the video game.

The Teacher Portal, as previously described, acts as a Content Management System (CMS) for the MGS. The Teacher Portal is accessible to participants via the internet, as a web page. The web page utilises PHP and jQuery to provide a consistent and modern user interface. Tooltips and hints are provided throughout the system, to promote correct and meaningful engagement with the system. Teacher Participants are allowed to modify game content but are not responsible for adding or removing students from the system. Students initially create their player profile in the MGS, which the teacher

can view, but cannot modify. The teacher is able to view specific game metrics, in both a tabular and graph form.

4.8 Summary

This Chapter has presented a detailed process for the design of a serious game for reading comprehension. This Chapter began by describing the requirements elicitation process with teachers from a Tasmanian school. Several of these requirements were then translated and used as input for the Behaviour Change Wheel Methodology. The result of this methodology was a detailed analysis of the problem behaviour identified, along with several target behaviours and strategies which could enable behaviour change. The link between game elements and Behaviour Change Techniques was then explored, with the result of this process being a map of game elements that were tied to learning mechanics. This Chapter then presented a comprehensive design summary of a Mini-Game Systems for reading comprehension. The Teacher Portal was described in detail, along with the design decisions and stylistic choices of the MiniNauts serious game. Reading comprehension activities were described in the manner that they appeared in the MGS, and systems for gamifying this content were explored. The following Chapter will describe the second portion of this study's methodology, with respect to data collection techniques, tools, and cohort selection.

5 Experiments

As mentioned in the preceding Chapters, studies in the serious games space are regularly empirical, and interventionist in nature (Connolly 2012b). The most common method for empirically measuring the outcomes of serious games, according to Connolly (2012a), is use the pre/post-test measurement tool. In the educational game space, pre/post-test measurements have been successfully used for testing university level computer science students (Eagle & Barnes 2009), high-school history students (Huizenga et al. 2009), and university level natural science students (Hwang et al. 2013). Both quasi-experimental, and randomised-control designs are employed in this research space. Due to the nature of internal validity being potentially compromised in quasi-experimental designs, Grimshaw et al. (2000) argue that randomizing participants into control/test cohorts, maximises the opportunity for causality to become the prime factor in any differences between the cohorts. This study follows the randomized control trial design, in an effort to reduce internal validity errors.

The intervention phase of this study involved deploying the complete system into a primary school and performing real-time measurements of student performance in the domains of reading comprehension, and general gameplay performance. The intervention was administered to a cohort as defined in section 5.1.This cohort was constrained by several factors. Due to time and resource constraints, a sample size (n=54) was chosen, with a generalizable population of 43,854 Tasmanian primary school students (Australian Bureau of Statistics 2010). As such, the generalisability of this study is partially limited, due to the difficult nature of sampling a large population size of students, when viewed from an ethical approval standpoint. The intervention ran for a period of 8 weeks, which conforms to one term length for a government school in Tasmania. The intervention is broken into four distinct components, each of which attempt to answer part of, or multiple research questions.

5.1 *Experimental Phase*

The purpose of this section is to describe the way the MGS was empirically tested in an educational setting. The study participants are split into multiple groups and exposed to different interventions of the system. Each intervention represents an independent variable, such that the Primary Study design represents a Mixed Methods design.

5.1.1 *Primary Study*

The procedure for the study is outlined in the following section. Figure 5-1 illustrates the specific procedure of an individual cohort. The following procedure is intended to provide a baseline for the individual student (and cohort) before the intervention begins (Pre-Test). Following this, the intervention for each group will be composed of a literacy task—detailed further—and concludes with a Post-Test evaluation. The intention of the Post-Test is to determine if significant changes in the learner's literacy has occurred after being exposed to the specific intervention. The purpose of the control group in is to determine a baseline for the between-group design, through which each intervention may be compared. The control group was composed of a random sample of participants, to ensure statistical comparability between the cohorts.

Figure 5-1 Experimental Procedure

5.1.1.1 Detailed Testing Procedure

The following section outlines the specific procedure that will be followed in the intervention.

1. Spend 1 hour instructing the teacher on the use of the software
2. All students in the class will be part of the study (if valid consent is given by each student).

3. At the teacher's discretion, the students are asked to complete one version of an online pre-test (Appendix A).
    a. The student navigates to a webpage and is asked to sign up for the first time. The student enters a username and a password, which will be re-used during the course of the study.
    b. The test involves reading a passage of text and answering via multiple choice questionnaires.
    c. Once the test is complete, the students logs off and continue classroom activities
4. Over the course of 8 weeks, students are asked by the teacher to play the game a minimum of three times per week. This process will be at the discretion of the teacher, and depend on computer availability, time constraints and teaching plans. The overall aim of the research is to present a piece of software that augments but does not disrupt normal teaching. The aim is for each student to be exposed to the game a minimum of 24 times in 8 weeks. Extant literature does not list a weekly measure for game administration, and this number was primarily chosen due to the constraints and needs of teachers, particularly limiting the study to the length of one school term.
    a. Each student logs on to a classroom PC and open a browser. The student then navigates to a website.
        a) Each student is then randomly assigned to a different experimental group. These groups break down the game into slightly different user experiences
            - Game with no feedback.
                - Students are not informed of any progress they have made over time or during the session. This data will still be recorded and sent to the server.
            - Game with feedback.
                - The game provides feedback to the students, as has been described in the preceding sections.
            - Game with no reading comprehension activities (Control).

- The game is thematically and mechanically the same as the previous two groups but will not include reading comprehension activities embedded into the game. The specific nature of the control games is outlined in Section 4.6.

b) The ways in which these differences will be apparent to students is via the end of game and between games screen, where their scores and progress are displayed, and any tests are administered. The information on who is in which group will not be presented to the students, as social competition is not a focus of this study. Some competition may occur between students if they share their results with their peers, but the knowledge of this is not expected to impact the results of the participant's data in a measurable manner.

b. The game is composed of two Mini-Games with feedback features at the end of the gameplay (dependant on which cohort the intervention is being administered to)

i. Students are informed to attempt as many mini-games as they like, in any order they like.

ii. Students will be able to retry the games as many times as they like.

iii. The system records in-game actions, their responses to any explicit gameplay decisions or questions and send this data in real-time back to an external server hosted at the University. This data is encrypted in transmission and storage.

iv. Once the game is complete or the student is finished playing, the player ends the game, and is presented with overall feedback on how they have performed in the current session, and how this compares to any previous sessions (TG3) (this data is retrieved from the server at login).

5. The system records the in-game actions of the players via screen recording software, for later analysis. This data will be stored on an encrypted server located at the University.
    a) Once the main study time period has been completed, the students will be asked by the teacher to engage in a post-test, and then answer a survey
    a. The post-test mirrors the pre-test in terms of question styles but differ in the actual passage text.
    b. The post-test procedure is identical to the pre-test procedure
    c. Students are asked at the end of the post-test to answer a Game Experience Questionnaire which details their feelings of enjoyment and satisfaction towards the game.
        i. Students will have completed the testing and gameplay portion of the study.
6. The Teacher will be invited to a debriefing session, where their experiences will be recorded via a semi structured interview
7. Each class of students will be interviewed about their experiences towards the game, using a post-study survey, developed in conjunction with the teacher of the class.

5.1.2 *Analysis Phase*

The Analysis phase of this study attempts to statistically, and critically present the data gathered in the Intervention phase. The methods of statistical analysis are described in detail in Chapter 6.

5.2 Participants

Participants in this study are defined as students in year 5 and 6 (age 10–12) in Tasmanian Primary Schools. Beyond consent to this study, there were no other selection factors for participants. As classrooms of minors are a unique research environment, the teacher can also be considered as a participant in this study. The teacher participants administer the content to students, acting as a gatekeeper, while also designing content

for the system. Students interacted with the game during computer time allocated for their class. Teachers continued their literacy plans as normal, with some modifications relating to the number of reading comprehension activities decreasing due to the change in teaching content ratios. Teachers reported that no significant extra time was spent per week on literacy activities, with the inclusion of the game system. Students were asked to perform their activity independently, and to not share results between themselves during their time with the system. As in most quantitative experimental designs, increasing the sample size better generalises the results to the wider population via external validity (Calder et al. 1982). Due to a lower sampling size, efforts to target specific student characteristics or demographics were not made. This does not limit the validity of the research, beyond not targeting more select groups of students i.e. male students for more targeted lines of inquiry. Efforts for analysis on demographic and characteristic bounds have been made in Chapter 6, and the relevant discussion for these analyses inform potential areas for future research; the generalisability of these analyses however is lower due to cohort sample size.

Whilst generalisability is difficult to establish in this study, the resulting data generated from the cohort participation is therefore considered 'general' and would aid in understanding broader factors of student literacy in Australian Primary Schools. Effectively, the data and results generated from this study could inform future research with respect to study and game design in the area of serious games and literacy.

5.2.1 *Participant Selection*

Participants were selected and placed into one of the three experimental cohorts (see Section 5.3 for a description of these cohorts). The selection process attempted to randomise the distribution of participants into cohorts, following the precedent of (Zebel et al. 2013) whereby participants of a similar study were placed within a between-group design using randomisation. This process attempts to reduce the impact of concentrating certain participants within the same cohort, to increase external validity and generalisability. Randomisation was achieved during the sign-up phase of the intervention. No information as to which experimental group participants were placed in was presented to the students. Participants were recruited from a semi-regional

primary school located in a semi-affluent area of Hobart, Tasmania. The school reports a higher than average socio-economic adjust score for student outcomes when compared to schools of a similar socio-economic makeup, as reported by MySchool, an Australian government school ranking service (ACARA 2018). Although not collected by the demographic survey, ABS survey data sates that Tasmanian access to at-home internet is 78% (ABS 2016a), and use of video games is 30% (ABS 2016b). The measure of whether access to the internet or use of video games acting as an influence on the result of this study was not explored, due to part of this research focusing on whether time spent playing video games recreationally would influence the outcome of a serious game led intervention. At home resources were not explored as a serious point of investigation, due to the limited scope of this research, however future research may focus on the potential disparate outcomes between students with at home resources and those without with regards to serious games for literacy.

5.3 Cohorts

This study seeks to examine the role that feedback, assessment, and the exposure to mini-games may have on learning, motivation and enjoyment for participants. As such, the study participants were split into multiple cohorts. The purpose of these cohorts is to provide comparable groups of participants through which statistical testing can be performed. To answer the research questions (SRQ1, SRQ2, SRQ3, and SRQ4), three sub cohorts, or Test Groups, as they are referred to hereon were defined. The purpose of each test group is to be exposed to the intervention in differing capacities. The test groups (denoted as TG) form the basis of the variable testing. The Test Groups are set out in manner presented in Table 26.

Test Group	Test Group Distinctions
TG1	Provide a baseline of performance through which the following two TGs can be compared to.
TG2	A group in which the full game is administered, but where no feedback on student performance is presented.

TG3	A group where the full game is administered, along with structured periods of feedback are presented to each student on an individual basis.

Table 26 Study Test Groups

The purpose of each group is to test a different condition of the intervention across a randomised sample of the total cohort. As can be observed in the preceding section, each group represents an independent variable which is being manipulated to determine several factors. Figure 5-2 describes visually the segmentation of participants into different cohorts.

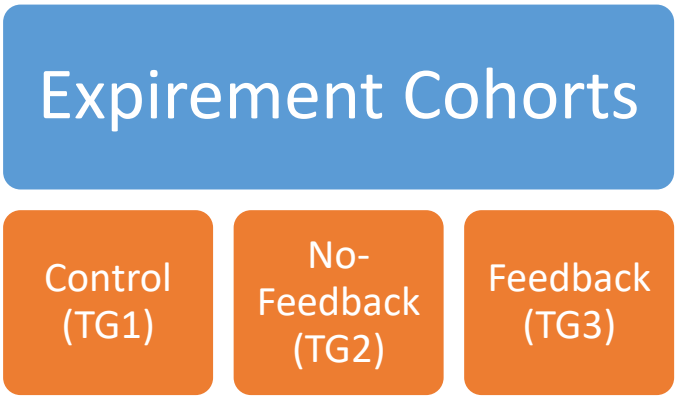


Figure 5-2 Cohort Design

The differences between each cohort, with respect to the game system and how features inter-relate is described in greater detail in Chapter 6.

5.4 *Data Collection*

A variety of data-points and metrics were captured by various assessments within and surrounding the intervention, following the information needs of the project defined in Section 3.2.5 The following data collection techniques follow the precedent of Zebel et al. (2013) for collecting relevant educational and motivational data.

5.4.1 *Pre/Post-Test and In-Game Metrics*

Following the conclusion of the intervention period, a post-test was administered to students. This post-test was identical in structure and format to the pre-test, with a set of 18 increasing difficult questions. The purpose of the post-test is to provide a second performance baseline of student reading comprehension levels. The

results of the difference between the pre-and post-tests informs the answer of SRQ1. Following the post-test, a Game Experience Questionnaire (GEQ) was administered to students. The purpose of the GEQ is to determine and define any experiential factors that arose during gameplay periods (IJsselsteijn et al. 2013). Further details on the GEQ are described in Section 3.10. The final component of this phase of the intervention involved surveying student opinions on the system, while also providing demographic information on a per-student basis. This data serves the dual purpose of determining subsets of students which responded better, or worse to the intervention, as well as providing students an opportunity for students to provide feedback on the game, system and intervention.

Feedback collection was concluded by surveying teacher opinions of the system. This feedback period involved an inductive semi-structured interview. The format of this interview is described in greater detail in section 3.4.1.3. The purpose of this phase of the intervention is to gather teacher opinions of the system and provide an outlet in which teachers can provide critical opinions on how well the system administered the content to students. This method of intervention debriefing forms a core component on user-centred design principles (Wood 1997).

A significant proportion of the data collected in this study was gathered through students and teachers interacting with the system. Each meaningful interaction with the system was recorded into a live database, described in Section 4.7.3.3. The purpose of in-game metrics is twofold: to gather meaningful data on how students interacted with the system, and to measure game, and reading performance as the participants played the game. As such, the in-game metrics inform the answering of all research questions. The use of in-game metrics to draw conclusions on player interaction correlational factors, is discussed in detail in Chapter 6.

5.5 *Ethical Considerations*

There are several key ethical considerations with respect to the experimental phase of this project. As the primary participants of this study are under the age of 18, they are considered minors as per Australian law (Health & Council 2017). The ethical guidelines for interacting with minors is far stricter than those for adults, with issues

such as mental well-being, mental development, social isolation and consent as foremost considerations. This study aims to respectfully integrate empirically tested intervention techniques into a classroom and provide students with an opportunity to improve their reading comprehension skill through increased motivation. Consent is provided by both the student and their parent/guardian prior to the commencement of their participation in this study. This consent can be withdrawn at any time and does not adversely affect the student in any meaningful way. The security of student data is a primary consideration with respect to the ethical concerns of this study. Student data is to be stored in an encrypted and inaccessible (beyond the primary research team) database, hosted at the University of Tasmania. The data used in this study will be re-identifiable, meeting the National Statement on Ethical Conduct of Human Research (4.2) guidelines for data security (ACARA 2017c), along with all guidelines from the Tasmanian Department of Education (Department of Education 2017). University ethics approval was granted under code H0015231.

6 Results

This Chapter will present the results of the intervention described in the preceding Chapters. This Chapter is organized into three main sections. Section 6.1 describes the results of the pre- and post-test administered to students across the three cohorts. The main aim of this section is to present data which is relevant to answering SRQ1 and SRQ2. The second and third sections of this Chapter, Section 6.2 and Section 6.3, are devoted to potential demographic covariates and motivational relationships respectively. These two sections aim to present data which will be used to answer SRQ3.

6.1 Pre- and Post-Testing

To determine the mean changes in variance within and between cohorts at the start and end of the intervention, several statistical tests were run to determine if significant variances lay between the three cohorts. The purpose of these tests is to inform SRQ1 and SRQ2.

6.1.1 *Within-Group Analysis*

To begin, three within group analyses of variances were performed across the three cohorts. These types of tests provide an insight into whether the mean scores for students changed between the two testing sessions, with each student score being counted as a repeated measure. As student scores are repeated measures, and only two measurements are being tested per participant, a paired-samples t-test was determined as appropriate for use in analysis. The following sections present the results of these tests.

6.1.1.1 Cohort 1

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the control group (TG1) pre- and post-test scores, normalized to a score out of 100. Data are mean ± standard deviation, unless otherwise stated; Table 27 and Figure 6-1 demonstrate these descriptive statistics.

		Mean	N	Std. Deviation	Std. Error Mean
Pair	Post-Test	50.917	18	22.8859	5.3943
	Pre-Test	24.378	18	10.1018	2.3810

Table 27 Group 1 Test Score Paired Samples Statistics

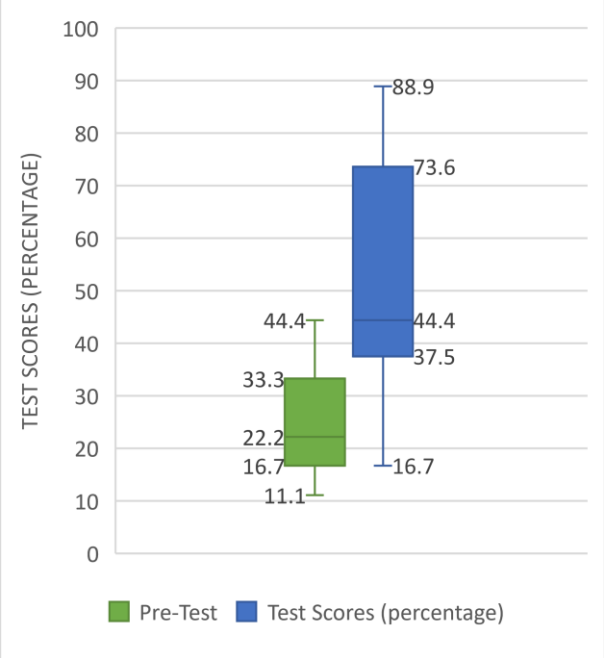


Figure 6-1 Group 1 Score Distribution

No outliers were detected. The assumption of normality was not violated, as assessed by Shapiro–Wilk's test (p = 0.599) presented in Table 28.

	Group	Kolmogorov–Smirnov			Shapiro–Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Difference	Group 1	.124	18	.200*	.960	18	.599

Table 28 Group 1 Test Score Tests of Normality

Participants performed higher overall in test scores in the post-test condition (50.915 ± 22.886) as opposed to the pre-test condition (24.378 ± 10.102), a statistically significant increase of (95% CI, 13.602 to 39.476) percent t(17) = 4.328, p < 0.0005, d = 1.020, with a power level of 94.8% reported, as shown in Table 29.

Paired Differences								
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tail)
				Lower	Upper			
Pre/Post	−26.539	26.016	6.132	−39.476	−13.602	−4.328	17	.000

Table 29 Group 1 Test Score Paired Samples Test

6.1.1.2 Cohort 2

A second paired-samples t-test was used to determine whether there was a statistically significant mean difference between the game with no feedback group (TG2) pre- and post-test scores, normalized to a score out of 100. Data are mean ± standard deviation, unless otherwise stated; Table 30 and Figure 6-2 present this data descriptively, and visually respectively.

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Post-Test	69.450	18	17.4913	4.1227
	Pre-Test	24.383	18	9.1633	2.1598

Table 30 Group 2 Test Score Paired Samples Statistics

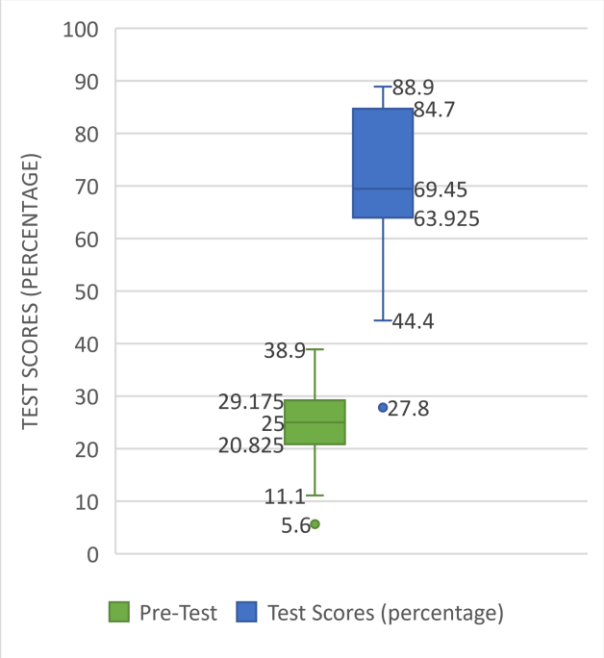


Figure 6-2 Group 2 Score Distribution

Six outliers were detected that were more than 1.5 box-lengths from the edge of the box in a boxplot. Inspection of their values did not reveal them to be extreme and they were kept in the analysis. The assumption of normality was not violated, as assessed by Shapiro–Wilk's test (p = 0.058) as presented in Table 31.

	Group	Kolmogorov–Smirnov			Shapiro–Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Difference	Group 2	.216	18	.025	.900	18	.058

Table 31 Group 2 Test Score Tests of Normality

Participants performed higher overall in test scores in the post-test condition (69.450 ± 17.491) as opposed to the pre-test condition (24.383 ± 9.163), a statistically significant increase of (95% CI, 35.885 to 54.248) percent t(17) = 10.356, p < 0.0005, d = 2.441 when examined through a paired samples T-Test, as shown in Table 32.

Paired Differences								
Tests	**Mean**	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Difference**		**t**	**df**	**Sig(2-tail)**
				Lower	**Upper**			
Pre/Post	45.0667	18.4633	4.3518	35.8851	54.2482	10.356	17	.000

Table 32 Group 2 Test Score Paired Samples Test

6.1.1.3 Cohort 3

A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the game with feedback group (TG3) pre-and post-test scores, normalized to a score out of 100. Data are mean ± standard deviation, unless otherwise stated; and data pertaining to these scores are presented in Table 33, and Figure 6-3.

		Mean	**N**	**Std. Deviation**	**Std. Error Mean**
Pair 1	Post	74.07	18	16.597	3.912
	Pre	28.71	18	11.467	2.703

Table 33 Paired Samples Statistics

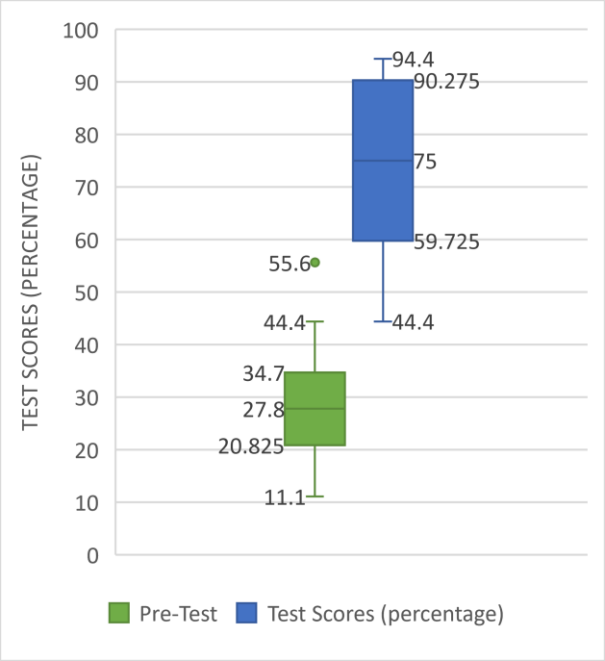


Figure 6-3 Group 3 Score Distribution

One outlier was detected that was more than 1.5 box-lengths from the edge of the box in a boxplot. Inspection of its values did not reveal them to be extreme and they were kept in the analysis. The assumption of normality was not violated, as assessed by Shapiro–Wilk's test (p = 0.227) shown in Table 34.

	Kolmogorov–Smirnov			**Shapiro–Wilk**		
	Statistic	df	Sig.	Statistic	df	Sig.
Difference	.161	18	.200*	.934	18	.227

Table 34 Group 3 Test Score Tests of Normality

Participants performed higher overall in test scores in the post-test condition (74.07 ± 16.597) as opposed to the pre-test condition (28.71 ± 11.467), a statistically significant increase of (95% CI, 37.600 to 53.133) percent t(17) = 12.324, p < 0.0005, d = 2.905, shown in Table 35.

Paired Differences								
Tests	**Mean**	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Difference**		**t**	**df**	**Sig(2-tail)**
				Lower	**Upper**			
Pre/Post	45.367	15.618	3.681	37.600	53.133	12.324	17	.000

Table 35 Group 3 Test Score Paired Samples Test

6.1.1.4 Within-Group Summary

To test the changes in variance between the TG1 pre- and post-test condition, a paired-samples t-test was run. The t-test measures variances between two cohorts across one independent variable, in this case, TG1 pre- and post-test. The result of this test demonstrated a statistically significant change in mean scores between pre- and post-test conditions for all three cohorts. This demonstrates that for students in TG1, TG2, and TG3 overall, a mean improvement across the 8-week intervention was realized. More simply, it can be asserted that student performance significantly improved on average, regardless of cohort. While a change in mean was detected for all three cohorts

in relative terms, effect size comparisons can only be made between the cohorts, to determine if the different cohorts outperformed each other. The following section describes the between group analysis of the three cohorts with respect to pre- and post-test scores.

6.1.2 *Pre-Test Between Group Analysis*

To determine whether pre-test scores were normally distributed between the three cohorts, a one-way ANOVA was conducted. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18). There were six outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro–Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .748$), shown in Table 36.

Levene Statistic	df1	df2	Sig.
.292	2	51	.748

Table 36 Homogeneity of Variances for Pre-Test Scores

Data is presented as mean ± standard deviation. Pre-test score increased from control (24.378 ± 10.102), no-feedback (24.383 ± 9.163), feedback (28.706 ± 11.467) groups, in that order, and the differences between these groups was considered not statistically significant, $F(2) = 1.060$, $p = .354$. This data is presented in Table 37.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	224.468	2	112.234	1.060	.354
Within Groups	5397.706	51	105.837		
Total	5622.173	53			

Table 37 One-Way ANOVA for Pre-Test Scores

The result of the test demonstrated no significant difference between cohort with respect to pre-test scores. This result allows the assertion to be made that the randomisation of student cohort placement successfully negated any biases between student performances. Therefore, the pre-test measure can act as a valid baseline for during conclusions between groups, as it is assumed that any changes in cohort scores

occurred after the administration of the pre-test. Descriptive statistics relating to the pre-test, factored by cohort is presented in Table 38.

Groups	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower	Upper		
Group 1	18	24.378	10.1018	2.3810	19.354	29.401	11.1	44.4
Group 2	18	24.383	9.1633	2.1598	19.827	28.940	5.6	38.9
Group 3	18	28.706	11.4674	2.7029	23.003	34.408	11.1	55.6
Total	54	25.822	10.2995	1.4016	23.011	28.633	5.6	55.6

Table 38 Descriptive Statistics for Pre-Test Scores

6.1.3 *Post-Test Between Group Analysis*

To determine the efficacy of the intervention with respect to test scores, and treating cohorts as an IV, a between group analysis is necessary to determine if variance lies between the cohorts. Further analysis with pairwise posthoc comparisons is presented where applicable to determine which pair contains significant variances in means.

A one-way ANOVA was conducted to determine if the post-test scores for each student was different between cohorts. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18). Further descriptive statistics are presented in Table 39.

Groups	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower	Upper		
Group 1	18	50.917	22.8859	5.3943	39.536	62.298	16.7	88.9
Group 2	18	69.450	17.4913	4.1227	60.752	78.148	27.8	88.9
Group 3	18	74.072	16.5968	3.9119	65.819	82.326	44.4	94.4
Total	54	64.813	21.3656	2.9075	58.981	70.645	16.7	94.4

Table 39 Descriptive Statistics for Post-Test Scores

There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro–Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .124$) as shown in Table 40.

Levene Statistic	df1	df2	Sig.
2.171	2	51	.124

Table 40 Post-Test Test of Homogeneity of Variances

Data is presented as mean ± standard deviation. Post-test score increased from control (50.917 ± 22.886), no-feedback (69.450 ± 17.491), feedback (74.072 ±16.597) groups, in that order, and the differences between these groups was considered statistically significant, $F(2) = 7.338$, $p = .002$ as shown in Table 41.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5406.175	2	2703.087	7.338	.002
Within Groups	18787.826	51	368.389		
Total	24194.001	53			

Table 41 One-Way ANOVA for Post-Test Scores

Bonferonni post hoc analysis revealed that the increase from control to no-feedback (18.533, 95% CI (2.695 to 34.371)) was statistically significant (p = .017), as well as the increase from control to feedback (23.156, 95% CI (7.318 to 38.993), p = .002), but the other pair of group differences was considered not statistically significant as shown in Table 42.

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower	Upper
Group 1	Group 2	−18.5333	6.3978	.017	−34.371	−2.695
	Group 3	−23.1556	6.3978	.002	−38.993	−7.318
Group 2	Group 1	18.5333	6.3978	.017	2.695	34.371
	Group 3	−4.6222	6.3978	1.000	−20.460	11.216
Group 3	Group 1	23.1556*	6.3978	.002	7.318	38.993
	Group 2	4.6222	6.3978	1.000	−11.216	20.460

Table 42 Bonferroni Pairwise Comparisons for Pre-Test Scores

Through the use of a one-way ANOVA, it was determined that there was a statistical difference between the cohorts, with respect to post-test score. Effect sizes and statistical power were calculated for both significant pairs of TG1-TG2, and TG1-TG3, and the non-significant pair of TG2-TG3. For TG1-TG2, an effect size for this analysis (d = .91, effect size = 0.41) and was found to exceed Cohen (1980)'s convention for a large effect size (d = .80). A power level of 0.67 was identified. For TG1-TG3, an effect size for this analysis was calculated (d = 1.16, effect size = .50) and was found to exceed a large effect size limit (d = .80). A power level of 0.83 was identified. For TG2-TG2, an effect

size for this analysis (d = .28, effect size = .27) and was found to not exceed Cohen (1980)'s convention for a large effect size (d = .80). A power level of 0.35 was identified. It was determined that mean post-test scores differed significantly between the control and no-feedback conditions, and control and feedback conditions. It was also determined that there was no statistically significant variance between the no-feedback condition, and the feedback condition. More simply, it can be stated that students who engaged with the system, performed significantly better with a gamified version of the reading content. These results provide data through which SRQ1 can be answered in the affirmative: there exists a positive difference in test-scores between the control, and no-feedback cohorts.

While the gamified version of the system appears to have influenced post-test scores to be higher—both with large effect sizes—the feedback condition did not provide a statistically meaningful difference between the two gamified cohorts. This leads to the assertion that while students may have been at an advantage with the gamified system, feedback mechanisms present it TG3 did not produce statistically higher scores. This data can then be used to answer SRQ2 in the negative: there was no benefit to students playing the feedback game condition, over the non-feedback game condition. This section has presented data analysis which informs SRQ1 and SRQ2. The following sections will present data in an attempt to answer SRQ3.

6.2 *Demographic and Enjoyment Interactions*

To determine whether any external factors acted as a covariate to the scores of students, correlational analysis was performed on several demographic and enjoyment factors. Where applicable, these correlations are then followed by analyses of covariance. Demographic factors such as age, gender, and activity interaction time are theorised to potentially play a role in how players enjoy the system.

6.2.1 *Gender*

Gender was considered as a key demographic factor through which variance in cohorts may be affected. Figure 6-4 visually describe the scores of males and females via

changes in scores between pre- and post-test, on an absolute scale. Analysis is then performed to determine if gender acts as a covariate to scores.

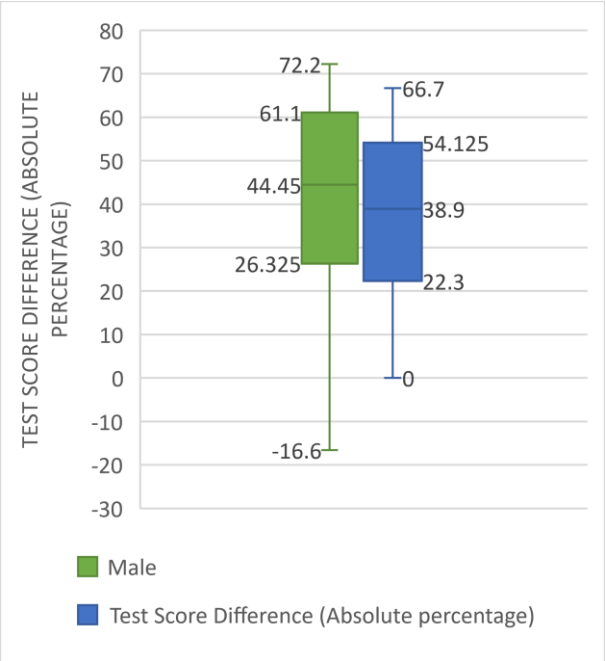


Figure 6-4 Test Score Change for All Cohorts against Gender

There were 30 male and 24 female participants, as shown in Table 43.

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Pre-Test	Male	30	24.447	11.3542	2.0730
	Female	24	27.542	8.7373	1.7835

Table 43 Pre-Test Group Statistics by Gender

An independent-samples t-test was run to determine if there were differences in pre-test scores between males and females. There were no outliers in the data, as assessed by inspection of a boxplot, presented in Figure 6-5.

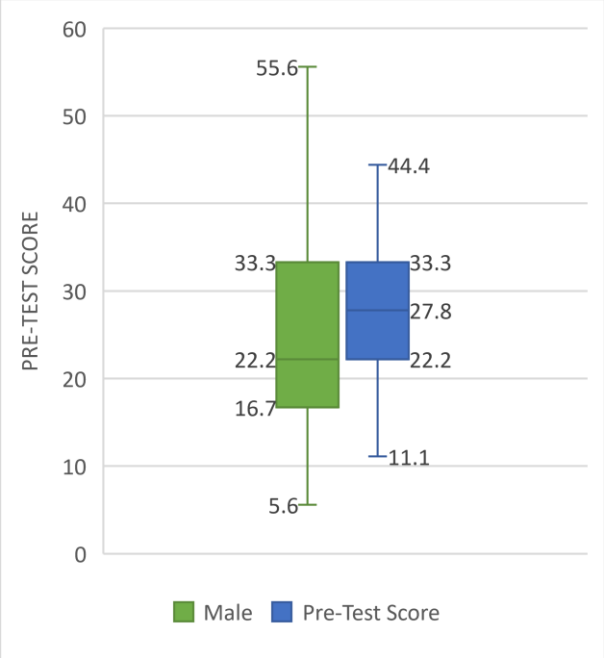


Figure 6-5 Pre-Test Distributions by Gender

Pre-test scores for each level of gender were normally distributed, as assessed by Shapiro–Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .137$). Pre-test scores were higher for females (27.542 ± 8.737) than male participants (24.447 ± 11.354), but no statistically significant difference existed 3.095 (95% CI, −8.744 to 2.5539), $t(52) = -1.009$, $p = 0.277$, as shown in Table 44.

	Levene's Equality of Variances								
	F	Sig.	t	df	Sig. (2-tail)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variance	2.285	.137	−1.099	52	.277	−3.0950	2.8151	−8.7439	2.5539
Unequal variances			−1.132	51.939	.263	−3.0950	2.7346	−8.5826	2.3926

Table 44 Independent Samples Test Pre-Test by Gender

An independent-samples t-test was run to determine if there were differences in post-test scores between males and females. There were no outliers in the data, as assessed by inspection of a boxplot presented in Figure 6-6.

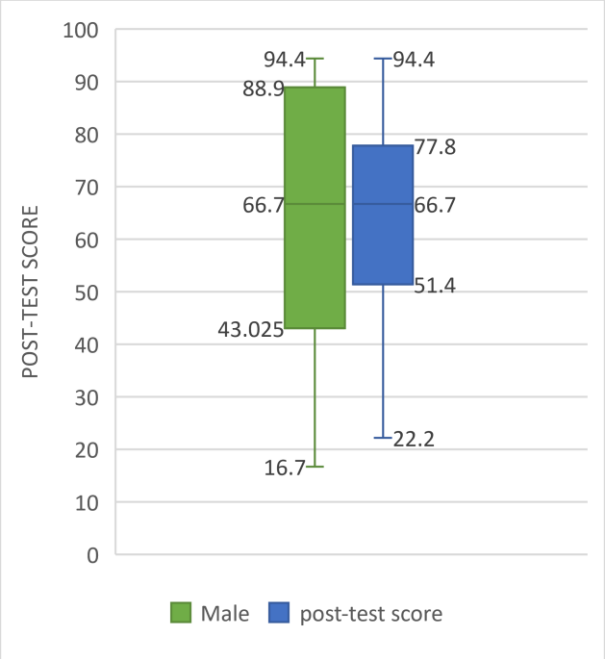


Figure 6-6 Post Test Distributions by Gender

Post-Test Scores were recorded as a mean of 64.073% for males and 65.738% for females, as shown in Table 45.

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Post-Test	Male	30	64.073	23.7362	4.3336
	Female	24	65.738	18.4371	3.7635

Table 45 Group Statistics Post-Test by Gender

Post-test scores for each level of gender were normally distributed, as assessed by Shapiro–Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .083$) shown in Table 46. Post-test scores were higher for females (65.738 ± 18.437) than male participants (64.073 ± 23.736), but no statistically significant difference existed 1.665 (95% CI, −13.509 to 10.1805), $t(52) = -0.282$, $p = 0.779$.

	Levene's Equality of Variances								
	F	Sig	t	df	Sig. (2-tail)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances	3.127	.083	−.282	52	.779	−1.6642	5.9027	−13.5088	10.1805
Unequal variances			−.290	51.968	.773	-1.6642	5.7397	−13.1818	9.8535

Table 46 Independent Samples Test Post-Test by Gender

Through categorizing participants into gender groups, and when comparing mean pre-test scores, it was determined that there was no statistically significant effect. It can therefore be asserted that gender did not influence pre-test scores, before the intervention in a significant manner. Categorizing participants into gender cohorts, it was determined through comparing mean post-test scores that no significant variance lay between genders with respect to post-test scores. It can therefore be asserted that gender did not act as a significant covariate between participants in the intervention, pre- and post-intervention.

6.2.2 *Age*

A one-way ANOVA was conducted to determine if significant differences exist between age groups with respect to pre-test scores. Participants were classified into three age groups: 10 (n = 3), 11 (n = 15) and 12 (n = 36). There were 3 outliers, as assessed

by boxplot; data was normally distributed for each group, as assessed by Shapiro–Wilk test ($p > .05$); and there was a violation of the homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .044$). Due to the proximity of the homogeneity of variance to significance, the one-way ANOVA was still carried out. Data is presented as mean ± standard deviation. Pre-test scores, presented as means and SD for age 10 (16.667 ± 5.550), age 11 (24.080 ± 6.534), and age 12 (27.311 ± 11.447), and the differences between these groups was considered not statistically significant, $F(2) = 1.832$, $p = 0.171$ as shown in Table 47.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	376.807	2	188.404	1.832	.171
Within Groups	5245.366	51	102.850		
Total	5622.173	53			

Table 47 One-Way ANOVA for Pre-Test Score and Age

A one-way ANOVA was conducted to determine if significant differences exist between age groups with respect to post-test scores. Participants were classified into three age groups: 10 (n = 3), 11 (n = 15) and 12 (n = 36). There were 3 outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro–Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .306$). Data is presented as mean ± standard deviation. Post-test scores, presented as means and SD for age 10 (72.233 ± 28.868), age 11 (56.660 ± 23.028), and age 12 (67.592 ± 19.758), and the differences between these groups was considered not statistically significant, $F(2) = 1.614$, $p = 0.209$ as shown in Table 48.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1440.211	2	720.105	1.614	.209
Within Groups	22753.790	51	446.153		
Total	24194.001	53			

Table 48 One-Way ANOVA for Post-Test Score and Age

Visual inspection of scatterplot with respect to the variables of post test score and age revealed a violation of the linear relationship assumption, and no monotonic

relationship exists. Therefore, further correlational analysis was deemed as not necessary.

Through categorizing participants by age, (10, 11, and 12), it was determined through statistical analysis that no significant differences between pre-test scores occurred between ages. This lack of significant interaction was also present for post-test scores. It can therefore be concluded that age did not act as a covariate to test scores and can be considered not a factor in mean changes in test scores.

6.2.3 *Time Spent Playing Video Games Recreationally*

A Welch Test of Equality of Means was conducted to determine if significant differences exist between groups with respect to hours played of video games per week. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18).

Levene Statistic	df1	df2	Sig.
2.629	2	51	.082

Table 49 Test of Homogeneity of Variances for Time Spent Playing Video Games

There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro–Wilk test ($p > .05$) as shown in Table 50; and there was a lack of homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = 0.082$) as shown in Table 49.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.139	18	.200	.915	18	.106
Group 2	.218	18	.024	.930	18	.193
Group 3	.175	18	.151	.893	18	.043

Table 50 Tests of Normality for Time Spent Playing Video Games

Data is presented as mean ± standard deviation. Hours played, presented as means and SD (3.11 ±2.447), no-feedback (2.61± 1.461), feedback (3.06 ± 2.711) groups, in

that order, and the differences between these groups was considered not statistically significant, df1 = 2, df2 = 31.336 as presented in Table 51.

	Statistic[a]	df1	df2	Sig.
Welch	.368	2	31.336	.695

a. Asymptotically F distributed.

Table 51 Welch Test of Equality of Means for Time Spent Playing Video Games

Visual inspection of scatterplot with respect to the variables of post-test score and hours played revealed a violation of the linear relationship assumption, and no monotonic relationship exists. Therefore, further correlational analysis is not necessary.

Through analysing categorizing participants by cohort, a one-way ANOVA was run to determine whether hours spent playing video games outside of class was significantly different between cohorts. This is to ensure that interest in video games was evenly distributed between cohorts and did not act as a covariate to mean changes in test scores. Through statistical analysis, it was determined that hours spent playing video games was not significantly different between cohorts. Through visual analysis of the spread of hours spent per participant, when all participants were compared together, it was determined that there was no correlation between hours spent playing video games at home, and post-test scores. It can therefore be stated that post-test scores appear not to be correlated to video game interaction outside of school.

6.2.4 *Interest in Video Games*

An independent samples Kruskal–Wallis test was conducted to determine if significant differences exist between groups with respect to interest in video games. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18).

Levene Statistic	df1	df2	Sig.
2.419	2	51	.099

Table 52 Test of Homogeneity of Variances for Interest in Video Games

There were no outliers, as assessed by boxplot; data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table

53; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .099) as shown in Table 52.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.243	18	.006	.837	18	.005
Group 2	.244	18	.006	.810	18	.002
Group 3	.328	18	.000	.710	18	.000

Table 53 Tests of Normality for Interest in Video games

A Kruskal–Wallis test was conducted to determine if there were differences in subjective opinions concerning the efficacy of the system over a traditional system scores between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot. Median scores were not statistically significantly different between the different cohort, $\chi^2(2) = 0.692$, p = .707 as shown in Table 54.

Independent-Samples Kruskal–Wallis Test			
N	**Test Statistic**	**Degrees of Freedom**	**Asymptotic Sig (2-tail)**
54	.692	2	.707

Table 54 Kruskal–Wallis Test for Interest in Video Games

Visual inspection of scatterplot with respect to the variables of post-test score and interest in video games revealed a violation of the linear relationship assumption, and no monotonic relationship exists. Therefore, further correlational analysis is not necessary.

Through observation of the distribution of player interest in video games, it was determined that no correlation existed between post-test score, and interest in video games. Therefore, it can be stated that for participants with a higher interest in video games overall, there was no relationship with better post-test scores.

6.2.5 *Enjoyment of Reading and Time Spent Reading*

To determine how participants enjoyed reading, the post-study survey asked participants to rate their agreement to the following statement, on a Likert 5-point scale of agreement:

I enjoy reading.

The purpose of rating this statement is to determine whether enjoyment of reading differed between cohorts. As the purpose of the intervention is to determine whether reading comprehension skills can be altered via gamified activities, the enjoyability of reading as a factor of play is an important consideration. The following scores represent the cohort means: control (65.28 ± 38.481), no-feedback (70.83 ± 27.453), and feedback (68.06 ± 23.957).

An independent samples Kruskal–Wallis test was conducted to determine if significant differences exist between groups with respect to enjoyment of reading. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18).

Levene Statistic	df1	df2	Sig.
3.375	2	51	.042

Table 55 Test of Homogeneity of Variances for Enjoyment of Reading

There were no outliers, as assessed by boxplot; data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 56; there was no homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p < 0.05$) as shown in Table 55.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.266	18	.001	.802	18	.002
Group 2	.245	18	.006	.833	18	.005
Group 3	.225	18	.016	.886	18	.033

Table 56 Test of Normality for Enjoyment of Reading

A Kruskal–Wallis test was conducted to determine if there were differences in subjective opinions concerning the efficacy of the system over a traditional system scores between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot. Median scores were not statistically significantly different between the different cohorts, $\chi^2(2) = 0.160$, $p = .923$ as shown in Table 57.

Independent-Samples Kruskal–Wallis Test			
N	**Test Statistic**	**Degrees of Freedom**	**Asymptotic Sig (2-tail)**
54	.160	2	.923

Table 57 Kruskal–Wallis Test for Interest in Reading

Through statistical analysis, it was determined that there were no significant differences between cohort medians with respect to enjoyment of reading. Participants reported above average enjoyment of reading across the three cohorts.

Participants were surveyed with respect to their reading preferences external to school activities. Participants were asked to respond to the following question:

How many hours do you read by yourself, or with your family per day?

Participants responded to this question, and the results are displayed in Table 58 and visually represented in Figure 6-7. It can be observed that males spent an average of 0.8 hours per day reading (SD = 6.93) and females spent an average of 1.29 hours per day reading (SD = 3.27).

	0hrs	**1hr**	**2hrs**	**3hrs**	**4hrs**	**Mean**	**SD**
Male	12	15	1	1	1	0.8	6.93
Female	6	8	7	3	0	1.29	3.27

Table 58 Hours Reading per Day Descriptive

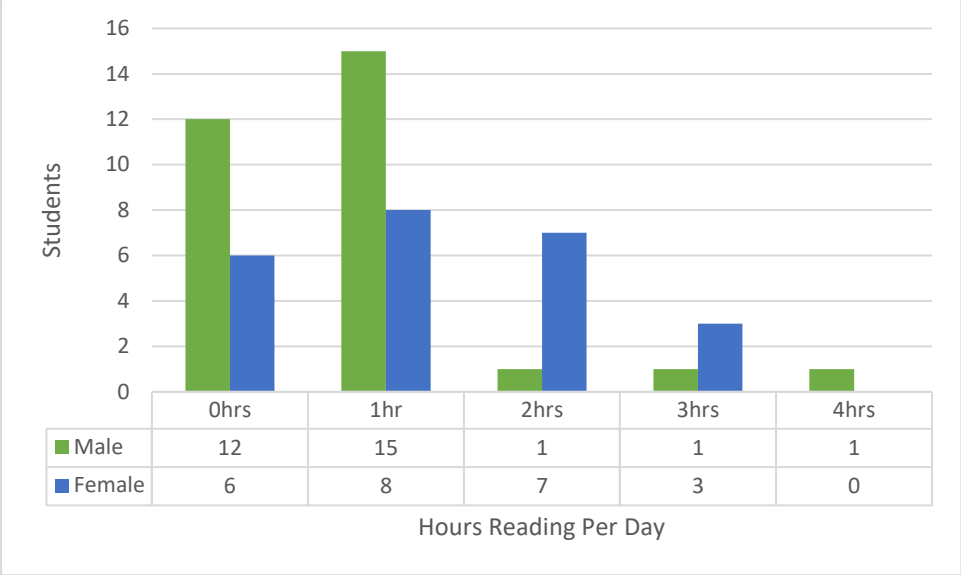


Figure 6-7 Hours Reading Per Day by Gender

6.2.6 *Video Game Preferences*

An intriguing secondary factor of serious game education is student's predilections for certain genres of video games. The post-study survey gathered information on the preferences of the two genders on their favourite video game genre. Preference information is demonstrated in Table 59.

Gender	Genre	Favourite Genre
Male	Sports	14
	Racing	9
	Action	4
	Puzzle/Building	2
	Adventure	1
	Social	0
Female	Puzzle/Building	12
	Sports	5
	Social	3
	Action	2
	Racing	2
	Adventure	1

Table 59 Favourite Genre by Gender

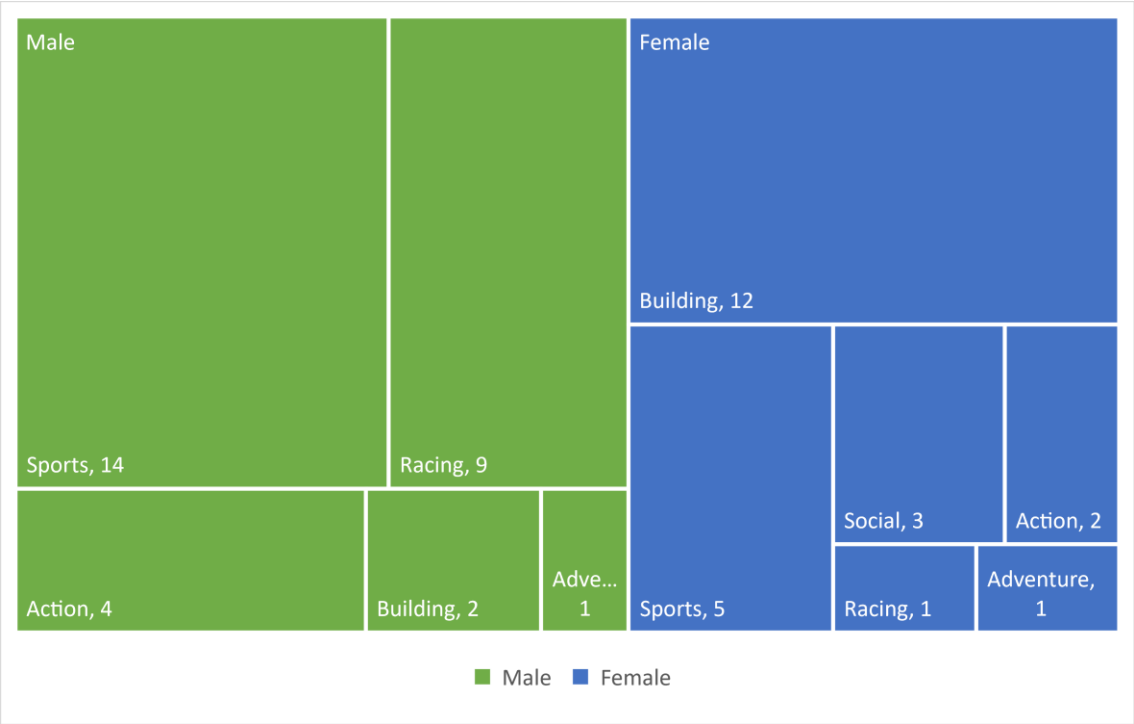


Figure 6-8 Favourite Genre by Gender Map

It can be observed that both genders shared dissimilar interests in video game genres. Males identified as preferring Sports, Racing and Action games as popular first genre choice. Female participants on the other hand preferred Building, Sports, and Social Games. A visual representation of this distribution is presented in Figure 6-8.

Data pertaining to the length of time participants spent playing video games recreationally per day was also collected. It was determined that males played an average of 3.8 hours (SD = 2.29) of video games per day. In contrast, females were identified as playing an average of 1.63 hours of video games per day (SD = 1.49); this comparison is represented visually in Figure 6-9.

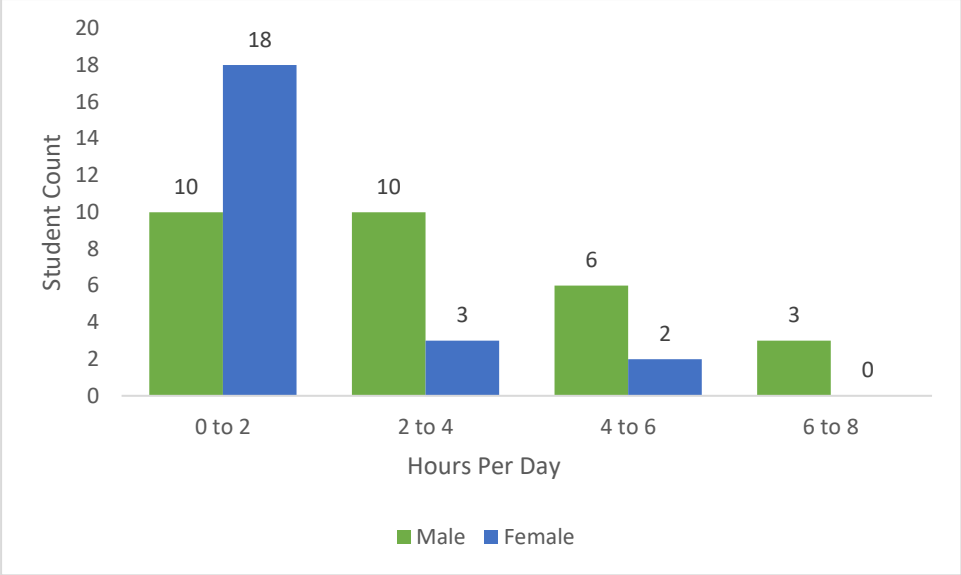


Figure 6-9 Average Daily Play-time by Gender

It can be stated that males self-identified as playing video games for longer than their female peers, by a factor of 2.33 times more hours per day, or a net difference of 2.17 hours per day.

Participants were surveyed regarding their preference to which device they most prefer to spend time playing video games on. Male and females shared a similar interest in devices, with Consoles and Smartphones being the two outliers of preference for males and females respectively. This information is presented visually in Figure 4-23.

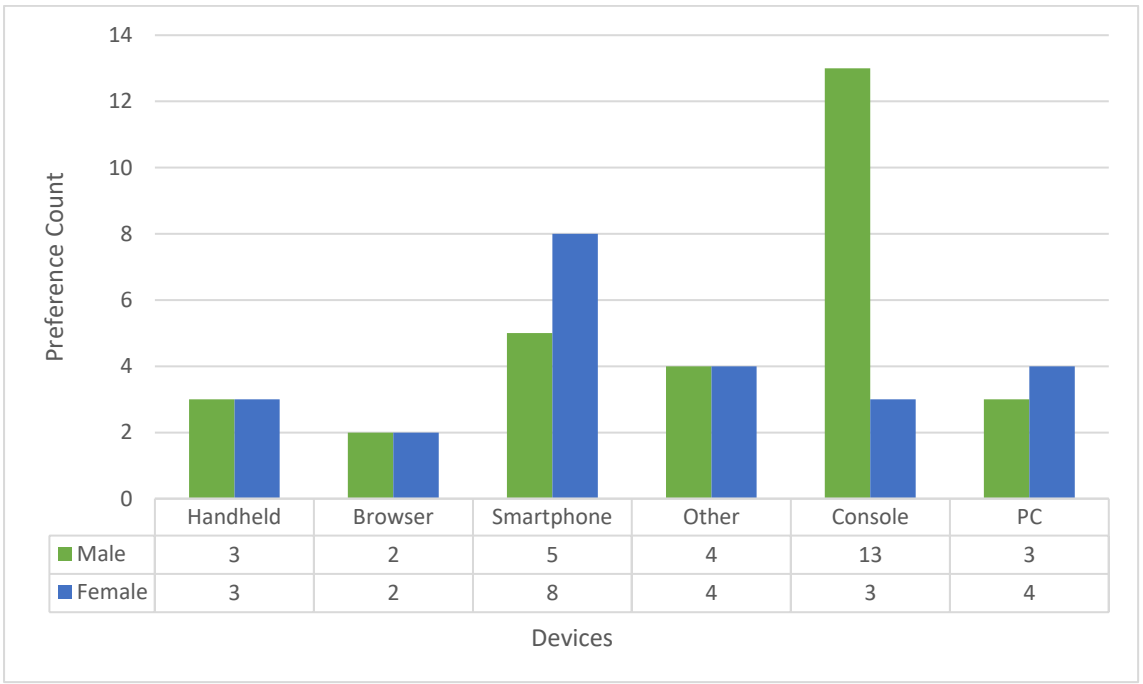
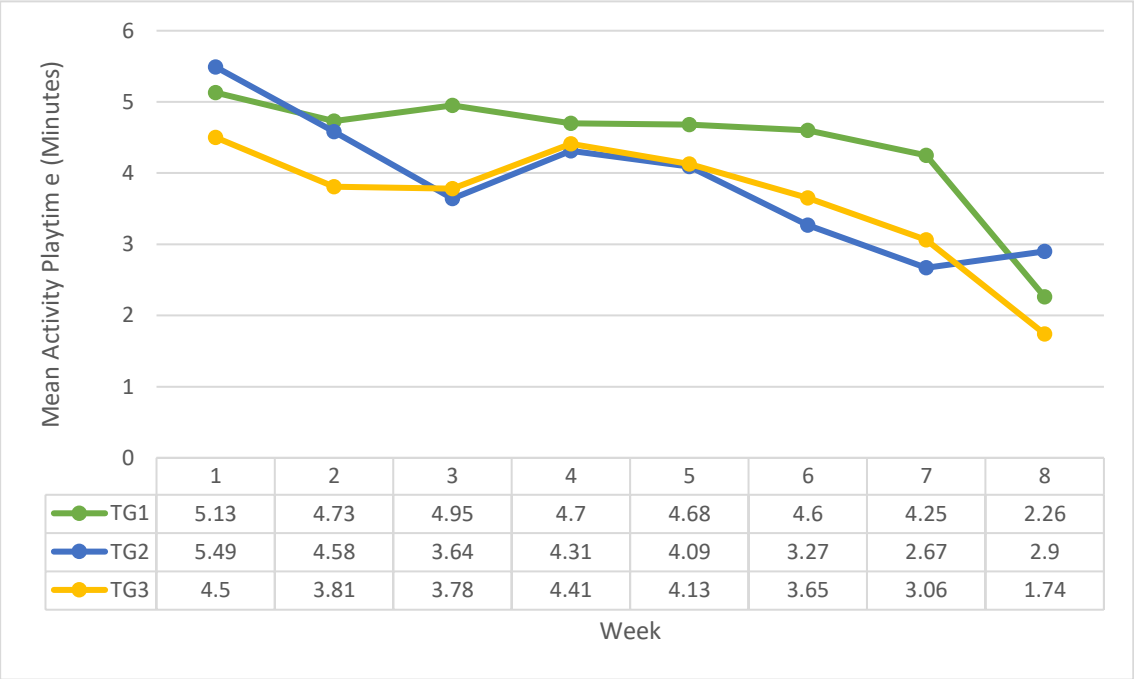


Figure 6-10 Device Preference for Video Games

6.2.7 *System Usage*

The MGS was deployed across two Tasmanian classrooms for a period of eight-weeks. The following section will discuss usage metrics of the system over time, and the manner in which students interacted with this system. Figure 6-11 lists the mean playtime per activity across the eight-week intervention. Note that the statistic refers to both Game 1 and Game 2 combined, excluding feedback periods for TG3. Activity time in Week 1 was consistently higher than activity time in Week 8 for all cohorts, from a combined mean of 5.04 minutes in Week 1, to 2.3 minutes in Week 8. Mean session times are represented in decimal minutes.



	1	2	3	4	5	6	7	8
TG1	5.13	4.73	4.95	4.7	4.68	4.6	4.25	2.26
TG2	5.49	4.58	3.64	4.31	4.09	3.27	2.67	2.9
TG3	4.5	3.81	3.78	4.41	4.13	3.65	3.06	1.74

Figure 6-11 Intervention Playtime by Test Group

Teachers generated 37 paragraph sections in the Teacher Portal and linked these paragraphs to 82 multiple-choice quizzes across the eight-week intervention. The mean Multiple-Choice Quiz (MCQ) to text ratio was 2.22 MCQs per paragraph. This equates to an average weekly reading comprehension activity number of 10.25 MCQs per week. The prescribed minimum activity days for this intervention was three days per week, leading to a mean activity count of 3.42 MCQs per activity day. In general terms, this meant that each participant played an average of 3.42 iterations of MiniNauts per

activity day. It was designed into the MGS for activities to be entirely voluntary and pursued in any order the student saw fit. Students were also able to revisit previous day's activities and complete them later. The mean completion rate for MiniNauts activities was 74.2%, bringing the mean activity completion to 2.53 MCQs per activity day. Between cohorts, it was identified that completion number amongst the TG1, TG2, and TG3 cohorts was 2.46, 2.28 and 2.76 minutes respectively. Readium, thinkium, and guessium scores were recorded across weeks, with respect to cohorts, these scores are presented in Figure 6-12. It was observed that TG3 had the highest mean of all three game elements.

	Guessium		Thinkium		Readium	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
TG1	2.04	0.48	2.94	0.75	2.18	0.69
TG2	2.12	0.25	2.81	0.63	2.31	0.45
TG3	2.40	0.51	2.92	0.73	2.83	0.44

Figure 6-12 Game Metrics by Cohort

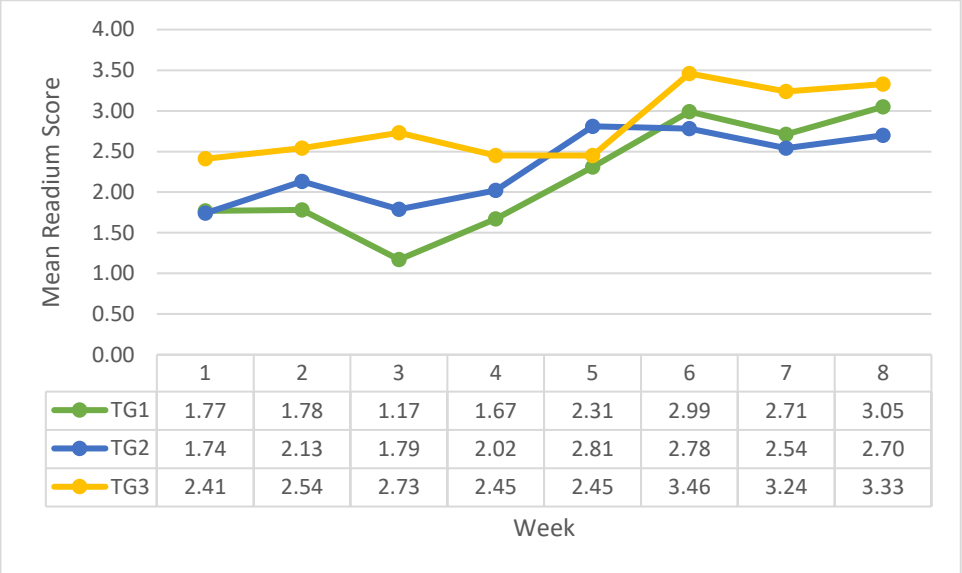
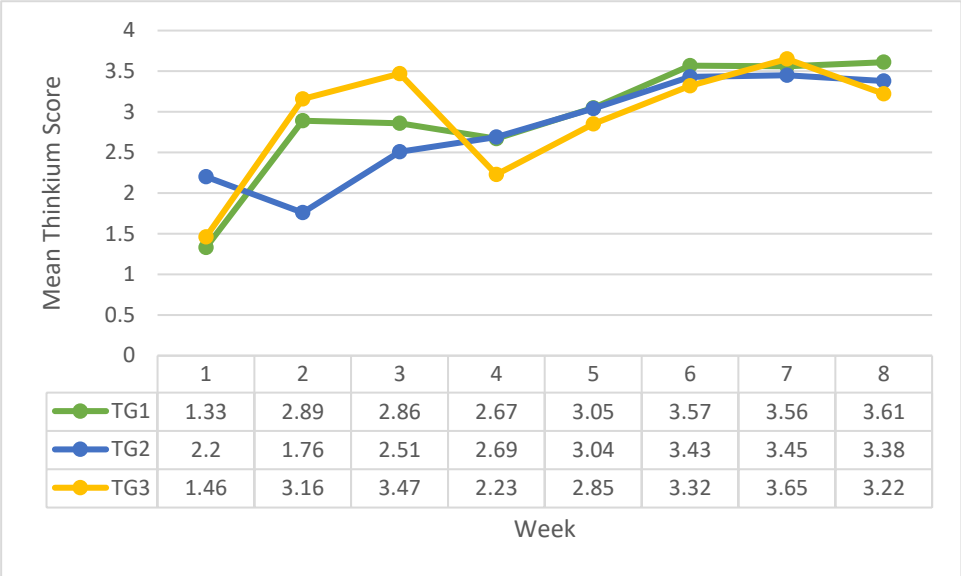
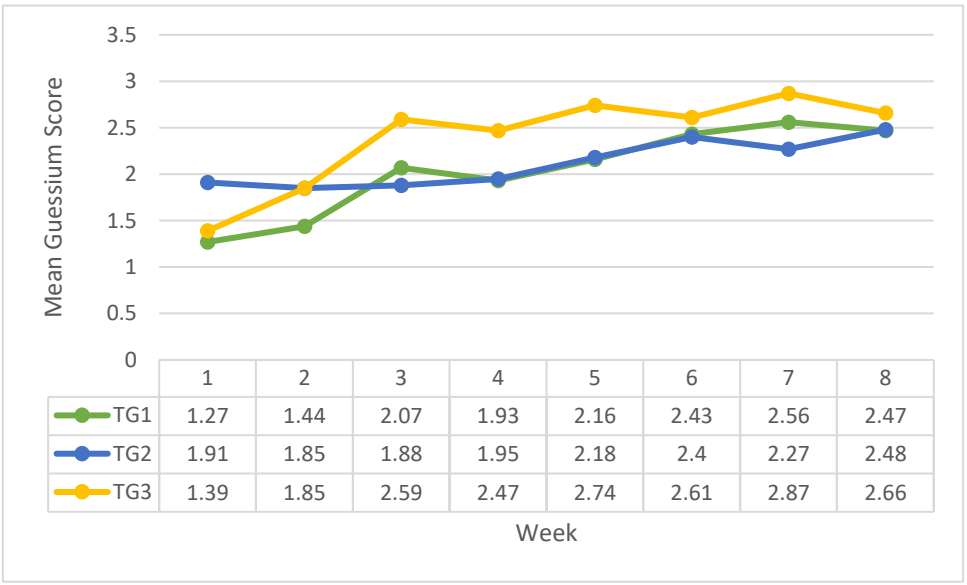


Figure 6-13 Readium Score by Week



	1	2	3	4	5	6	7	8
TG1	1.33	2.89	2.86	2.67	3.05	3.57	3.56	3.61
TG2	2.2	1.76	2.51	2.69	3.04	3.43	3.45	3.38
TG3	1.46	3.16	3.47	2.23	2.85	3.32	3.65	3.22

Figure 6-14 Thinkium Score by Week



	1	2	3	4	5	6	7	8
TG1	1.27	1.44	2.07	1.93	2.16	2.43	2.56	2.47
TG2	1.91	1.85	1.88	1.95	2.18	2.4	2.27	2.48
TG3	1.39	1.85	2.59	2.47	2.74	2.61	2.87	2.66

Figure 6-15 Guessium Score by Week

As can be observed in Figure 6-13, readium scores increased for all three cohorts across the eight-week intervention. This trend is continued for both thinkium (Figure 6-14) and guessium (Figure 6-15). Recall that readium measured player's ability to answer MCQs correctly; thinkium to infer meaning from sentences, and finally guessium to perform word substitution correctly. Scores for readium and thinkium were constrained to a maximum of four, per game session, with guessium constrained to three per game session. Teachers were capable of generating greater or fewer numbers of questions, and responses, and the MGS adjusted the weighting of each crystal to ensure the local maximum scores were maintained.

6.2.8 *Demographic and Enjoyment Summary*

This section presented several analytical tests with the purpose of determining if any external factors contributed to the acquisition of post-test scores. When analysing the demographic factors of age, and gender, it was determined that neither acted as covariates to the post-test scores of any cohort. For the factors of enjoyment with respect to video games, and reading, it was determined that there were no significant differences between cohorts. Briefly, this result is indicative that the intervention did not have the consequence of altering the enjoyment of video games or reading. Interest in video games as an activity were evenly distributed between the three cohorts, and no significant difference between the cohorts was identified. Students identified as playing different game genres by gender, with Sports and Racing games most favoured by males, and Puzzle/Building games most favoured by female participants. Males spent an average of 2.17 hours per day longer than females playing video games. Males were also identified as holding stronger interest for video games on consoles, when compared to females; smartphones were majorly preferred by females when compared to males.

6.3 Motivational and Enjoyment Factors

To determine whether relationships existed between different dependent variables in this study, several correlational tests were performed. These tests aim to expose meaningful relationships between enjoyability and motivational factors with respect to playing MiniNauts during the intervention.

6.3.1 *Usefulness of Feedback Features*

Feedback features form an important part of this study. As previously mentioned, feedback features for participants in TG3 provided feedback on past and current performance, along with information as to their progress. The post-intervention survey asked participants to rate the following statement from one to five on a five-point Likert scale of agreement:

MiniNaut's feedback and hints were useful for me playing the game.

The purpose of this question is to determine whether participants interacting with the game experienced differences in the usefulness of feedback features between cohorts. A one-way ANOVA was conducted to determine if the subjective usefulness of participants viewing their score was different between the three cohorts. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18). There were no outliers, as assessed by boxplot; data was normally distributed for Cohort 2 ($p > .05$), but was not normally distributed in cohort 1 (p = 0.02) and cohort 3 (p = 0.002) as assessed by Shapiro–Wilk test as shown in Table 61; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .0.061$) as shown in Table 60.

Levene Statistic	df1	df2	Sig.
2.956	2	51	.061

Table 60 Test of Homogeneity of Variances for Usefulness of Feedback Features

Due to the homogeneity factor, an ANOVA was still considered a valid test. Data is presented as mean ± standard deviation.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.216	18	.026	.873	18	.020
Group 2	.180	18	.129	.917	18	.113
Group 3	.214	18	.029	.812	18	.002

Table 61 Tests of Normality for usefulness of Feedback Features

Perception of the usefulness of feedback increased from control (47.22 ± 22.506), no-feedback (51.39 ± 32.621), feedback (76.39 ±20.059) groups, in that order, and the differences between these groups was considered statistically significant, $F(2) = 6.811$, $p = .002$, as shown in Table 62.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8958.333	2	4479.167	6.811	.002
Within Groups	33541.667	51	657.680		
Total	42500.000	53			

Table 62 One-Way ANOVA for Usefulness of Feedback Features

Bonferroni post hoc analysis revealed that the increase from control to no-feedback (4.167, 95% CI (17.00 to 25.33), (p = 1.00)) was not statistically significant, but the increase from control to feedback (4.167, 95% CI (17.00 to 25.33), p = 0.015) was significant, along with the increase from non-feedback to feedback (25.00, 95% CI (3.84 to 46.16), p = 0.015) as shown in Table 63.

(I) Group	(J) Group	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval	
					Lower	Upper
Group 1	**Group 2**	−4.167	8.548	1.000	−25.33	17.00
	Group 3	−29.167	8.548	.004	−50.33	−8.00
Group 2	**Group 1**	4.167	8.548	1.000	−17.00	25.33
	Group 3	−25.000	8.548	.015	−46.16	−3.84
Group 3	**Group 1**	29.167	8.548	.004	8.00	50.33
	Group 2	25.000	8.548	.015	3.84	46.16

Table 63 Bonferroni Pairwise Comparisons for Usefulness of Feedback Features

6.3.2 *Enjoyability of Feedback Features*

To determine whether participants enjoyed feedback features as part of the intervention experience, the post-intervention survey asked participants to rate the following statement from one to five on a five-point Likert scale of agreement:

MiniNaut's feedback and hints were fun and enjoyable.

The purpose of this question is to determine whether participants in different cohorts experienced the feedback features differently. A one-way ANOVA was

conducted to determine if the subjective usefulness of participants viewing their score was different between the three cohorts. Participants were classified into three groups: control (n = 18), no-feedback (n = 18) and feedback (n = 18).

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.180	18	.129	.917	18	.113
Group 2	.201	18	.054	.909	18	.082
Group 3	.187	18	.098	.908	18	.079

Table 64 Tests of Normality for Enjoyability of Feedback Features

There were no outliers, as assessed by boxplot; data was normally distributed for all cohorts ($p > .05$), as shown in Table 64; however, there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .061$) as shown in Table 65.

Levene Statistic	df1	df2	Sig.
.127	2	51	.881

Table 65 Test of Homogeneity of Variances for Enjoyability of Feedback Features

Due to the homogeneity factor, and distribution of normality, an ANOVA was deemed an acceptable statistical test. Data is presented as mean ± standard deviation.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	856.481	2	428.241	.405	.669
Within Groups	53958.333	51	1058.007		
Total	54814.815	53			

Table 66 One-Way ANOVA for Enjoyability of Feedback Features

Perception of the enjoyability of feedback: control (51.39 ± 32.621), no-feedback (56.94 ± 30.685), feedback (47.22 ± 34.180) groups, and the differences between these groups was not considered statistically significant, $F(2) = 0.405$, $p = 0.669$, as shown in Table 65.

6.3.3 *Feeling of Improvement*

An important consideration of to this study are factors of motivation and enjoyment. The post-study survey asked the following question of participants, and asked them to rank the agreement on a five-point Likert scale:

I got better at understanding what I read due to playing MiniNauts.

A Kruskal–Wallis test was conducted to determine if there were differences in feelings of personal improvement of reading skills, between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.296	18	.000	.856	18	.011
Group 2	.235	18	.010	.871	18	.019
Group 3	.225	18	.017	.814	18	.002

Table 67 Tests of Normality for Feeling of Improvement

Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 67; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .458$) as shown in Table 68.

Levene Statistic	df1	df2	Sig.
.793	2	51	.458

Table 68 Test of Homogeneity of Variances for Feeling of Improvement

Median scores were not considered statistically significantly different between the different cohorts, $\chi^2(2) = 3.356$, $p = 0.187$. Mean cohort rankings of improvement are as follows: control (38.89 ± 23.044), no-feedback (43.06 ±22.37), and feedback (52.78 ± 18.96), as shown in Table 69.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	3.356	2	.187

Table 69 Independent-Samples Kruskal–Wallis Test for Feeling of Improvement

6.3.4 *Game Difficulty*

To determine how participants felt about the game difficulty, the post-study survey presented the following statement to participants, and asked them to rank the agreement on a five-point Likert scale:

MiniNauts was hard to play.

A Kruskal–Wallis test was conducted to determine if there were differences in game difficulty, between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.218	18	.023	.827	18	.004
Group 2	.283	18	.001	.759	18	.000
Group 3	.308	18	.000	.786	18	.001

Table 70 Tests of Normality for Game Difficulty

Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 70; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .458$) as shown in Table 71.

Levene Statistic	df1	df2	Sig.
.163	2	51	.850

Table 71 Test of Homogeneity of Variances for Game Difficulty

Median scores were not considered statistically significantly different between the different cohorts, $\chi^2(2) = 0.649$, $p = 0.723$. Mean cohort rankings of game difficulty are as follows: control (26.39 ± 27.749), no-feedback (27.78 ± 22.506), and feedback (22.22 ± 25.565) as shown in Table 72.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	.649	2	.723

Table 72 Independent-Samples Kruskal–Wallis Test for Game Difficulty

Through statistical analysis, it was determined that participants in different cohorts did not significantly vary in their appraisal of game difficulty. When comparing means, it can be seen that the majority of participants moderately disagreed with the statement as to MiniNauts being difficult.

6.3.5 *Reading Difficulty*

An important consideration of this study is to determine the difficulty of reading activities for participants. The post-study survey asked the following question of participants, and asked them to rank the agreement on a five-point Likert scale:

The reading in MiniNauts was difficult.

A Kruskal–Wallis test was conducted to determine if there were differences in subjective opinions concerning the difficulty of reading between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.275	18	.001	.869	18	.017
Group 2	.322	18	.000	.705	18	.000
Group 3	.240	18	.008	.782	18	.001

Table 73 Tests of Normality for Reading Difficulty

Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 73; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .335$) as shown in Table 74.

Levene Statistic	df1	df2	Sig.
1.117	2	51	.335

Table 74 Test of Homogeneity of Variances for Reading Difficulty

Median scores were statistically significantly different between the different cohorts, $\chi^2(2) = 8.71$, $p < .013$. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values are presented. This post hoc analysis revealed statistically significant

differences in median scores between control (35.53) and feedback (21.11) ($p = 0.011$) as shown in Table 75; there were no significant differences between no-feedback and feedback conditions; and no difference between control, and no-feedback conditions.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	8.710	2	.013

Table 75 Independent-Samples Kruskal–Wallis Test for Reading Difficulty

Through statistical analysis, it was determined that participants in the feedback cohort, were significantly less likely to rate the reading portion of MiniNauts as difficult ($p < 0.05$), as shown in Table 76.

Samples	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
Group 3 – Group 2	4.750	4.979	.954	.340	1.000
Group 3 – Group 1	14.417	4.979	2.896	.004	.011
Group 2 – Group 1	9.667	4.979	1.942	.052	.157

Table 76 Kruskal–Wallis Pairwise Comparisons for Reading Difficulty

There were no other significant relationships identified via statistical analysis. With respect to mean ranking comparisons, it can be observed that most students rated the difficulty of reading in MiniNauts as either somewhat easy, or neutral.

6.3.6 *Game as a Substitute*

An important consideration of this study is to determine whether participants viewed the game as a better or worse substitute for reading activities; the post-study survey presented the following statement to participants, and asked them to rank the agreement on a five-point Likert scale:

MiniNauts is a better way to spend time reading than doing other reading activities.

A Kruskal–Wallis test was conducted to determine if there were differences in subjective opinions concerning the efficacy of the system over a traditional system scores between groups; shown in Table 79.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.260	18	.002	.807	18	.002
Group 2	.324	18	.000	.751	18	.000
Group 3	.248	18	.005	.866	18	.016

Table 77 Tests of Normality for Game Substitution Viability

Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot. Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 77; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .715$) as shown in Table 78.

Levene Statistic	df1	df2	Sig.
.338	2	51	.715

Table 78 Test of Homogeneity of Variances for Game Substitution Viability

Median scores were statistically significantly different between the different cohorts, $\chi^2(2) = 15.946$, $p < .0005$. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted *p*-values are presented. This post hoc analysis revealed statistically significant differences in median scores between control (16.42) and no-feedback (30.78) ($p = 0.011$), and control and feedback (35.31) groups ($p < 0.0005$), in pairwise combinations. There were no significant differences between no-feedback and feedback conditions.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	15.946	2	.000

Table 79 Independent-Samples Kruskal–Wallis Test for Game Substitution

Through statistical analysis using the Kruskal–Wallis independent samples test, it was determined that participants in TG2 ($p < 0.05$) and TG3 ($p < 0.05$) rated the experience of using MiniNauts as a substitute for other reading activities significantly higher than for participants in TG1, as shown in Table 80.

Samples	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.

Group 3 – Group 2	−14.361	4.939	−2.908	.004	.011
Group 3 – Group 1	−18.889	4.939	−3.824	.000	.000
Group 2 – Group 1	−4.528	4.939	−.917	.359	1.000

Table 80 Kruskal–Wallis Pairwise Comparisons for Game Substitution

Participants in TG2 and TG3 did not differ significantly in their assessment of the game as a substitute. More simply, it can be stated that students using a gamified version of reading activities identified the game as a better substitute for reading than students who did not interact with a game substitute.

6.3.7 *Student Effort*

To determine whether participants engaged with MiniNauts at different rates of self-identified effort, the post-intervention survey asked participants to rate the following statement from one to five on a five-point Likert scale of agreement:

I tried hard, and put in lots of effort into reading when playing MiniNauts.

The purpose of this question is to determine whether participants significantly differed in the amount of effort they put into playing MiniNauts. This metric is self-identified, and differs from post-test score, which measures absolute achievement, it does not measure the amount of effort put into playing.

A Kruskal–Wallis test was conducted to determine if there were differences in player efforts between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot.

Group	**Kolmogorov–Smirnov**			**Shapiro–Wilk**		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Group 1	.286	18	.000	.819	18	.003
Group 2	.211	18	.033	.883	18	.030
Group 3	.185	18	.103	.872	18	.019

Table 81 Tests of Normality for Student Effort

Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 81; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .606$) as shown in Table 82.

Levene Statistic	df1	df2	Sig.
.506	2	51	.606

Table 82 Test of Homogeneity of Variances for Student Effort

Median scores were not statistically significantly different between the different cohorts, $\chi^2(2) = 0.053$, $p = 0.974$. Through statistical analysis, it was determined that the three cohorts did not meaningfully differ from one another with respect to median scoring of effort. Participants rated the effort put into playing MinNauts as control (33.33 ± 27.166), no-feedback (37.5 ± 31.213), and feedback (36.11 ± 34.537) respectively as shown in Table 83.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	0.053	2	.974

Table 83 Independent-Samples Kruskal–Wallis Test for Student Effort

Participants across the three cohorts rated the effort they put into reading as moderately easy on average. It can therefore be stated that most participants did not feel that they put in a serious amount of effort into reading during the game.

6.3.8 *Enjoyability*

To determine the enjoyability of MiniNauts, the post-intervention survey asked participants to rate the following statement from one to five on a five-point Likert scale of agreement:

I found MiniNauts fun and enjoyable

The purpose of this question is to determine how participants felt MiniNaut's measured with respect to subjective enjoyment of the game. Participants responded to the question of enjoyability: control (51.39 ± 24.96), no-feedback (63.89 ± 23.044), and feedback (62.5 ±17.68) respectively. It can therefore be stated that many participants found MiniNauts moderately enjoyable, across all three cohorts.

A Kruskal–Wallis test was conducted to determine if there were differences in player determinations of fun between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.311	18	.000	.853	18	.009
Group 2	.227	18	.015	.889	18	.037
Group 3	.273	17	.001	.848	17	.010

Table 84 Tests of Normality for Enjoyability

Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 84; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .612$) as shown in Table 85.

Levene Statistic	df1	df2	Sig.
.495	2	51	.612

Table 85 Test of Homogeneity of Variances for Enjoyability

There were no outliers; Median scores were not statistically significantly different between the different cohorts, $\chi^2(2) = 2.730$, $p = 0.255$ as shown in Table 86.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	2.730	2	.255

Table 86 Independent-Samples Kruskal–Wallis Test for Enjoyability

Through statistical analysis, it was determined that participants in all three cohorts, did not differ significantly in their enjoyment of the game across medians subjective rankings. More simply, it can be stated that participants enjoyed the game moderately, across all three cohorts.

6.3.9 *Interest in Video Games with respect to Enjoyment of MiniNauts*

To determine whether participants with an interest in video games had an increased level of subjective enjoyment of MiniNauts, correlational analysis was run.

The importance of this relationship is to determine whether participants who enjoy video games outside of school enjoyed the game more than students who identify as not being as interested in games. The measurement of interest in games was determined through asking participants to rank their agreement on a five-point Likert scale for the following statement:

I generally enjoy playing video games. (Either games for fun, or educational games).

A Spearman's rank-order correlation was run to assess the relationship between interest in video games and enjoyment of the MiniNauts game. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. Data was determined to be not normally distributed ($p < 0.05$) by a Shapiro–Wilk test of normality, shown in Table 87.

Kolmogorov–Smirnov			**Shapiro–Wilk**		
Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
.272	54	.000	.795	54	.000
.234	54	.000	.873	54	.000

Table 87 Tests of Normality for Interest in Video Games

Through analysis using a Spearman Correlation Test, it was determined that there was no correlation between interest in video games, and enjoyment of MiniNauts, $r_s(54) = 0.104$, $p = 0.456$ as shown in Table 88.

			Game Enjoyment	Interest in Games
Spearman's rho	**Game Enjoyment**	**Correlation Coefficient**	1.000	.104
		Sig. (2-tail)	.	.456
		N	54	54
	Interest in Games	**Correlation Coefficient**	.104	1.000
		Sig. (2-tailed)	.456	.
		N	54	54

Table 88 Spearman Correlation Test for Interest in Video Games

Through correlational analysis, it was determined that there was no correlation identified between participants enjoying games, and participants enjoying MiniNauts during the intervention.

6.3.10 *Intrinsic Motivation*

To determine whether participants were motivated by the MGS, the post-study survey presented the following statement of participants, and asked them to rank the agreement on a five-point Likert scale:

I felt motivated by playing MiniNauts.

The purpose of this question is to determine whether participants in the three cohorts had a different subjective experience of motivation, when comparing mean subjective rankings. Motivation means for the three cohorts were: control (36.11 ± 23.044), no-feedback (58.33 ± 19.174), and feedback (69.44 ± 20.211) conditions respectively.

A Kruskal–Wallis test was conducted to determine if there were differences in subjective feelings of motivation that MiniNauts provided between groups. Distributions of scores were similar for all groups, as assessed by visual inspection of a boxplot.

Group	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group 1	.227	18	.015	.889	18	.037
Group 2	.279	18	.001	.863	18	.013
Group 3	.275	18	.001	.869	18	.017

Table 89 Tests of Normality for Intrinsic Motivation

Data was not normally distributed for each group, as assessed by Shapiro–Wilk test ($p < .05$) as shown in Table 89; there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .528$) as shown in Table 90.

Levene Statistic	df1	df2	Sig.
.646	2	51	.528

Table 90 Test of Homogeneity of Variances for Intrinsic Motivation

Median scores were statistically significantly different between the different cohort, $\chi^2(2) = 16.419$, $p < .0005$. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted *p*-values are presented and shown in Table 91.

N	Test Statistic	Degrees of Freedom	Asymptotic Sig (2-tail)
54	16.419	2	.000

Table 91 Independent-Samples Kruskal–Wallis Test for Intrinsic Motivation

This post hoc analysis revealed statistically significant differences in median scores of motivations between control (16.61) and no-feedback (29.31) ($p < 0.0005$), and control and feedback (36.58) groups ($p < 0.0005$), in pairwise combinations. There were no significant differences between no-feedback and feedback conditions as shown in Table 92.

Samples	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
Group 3 – Group 2	−12.694	4.989	−2.544	.011	.033
Group 3 – Group 1	−19.972	4.989	−4.003	.000	.000
Group 2 – Group 1	−7.278	4.989	−1.459	.145	.434

Table 92 Kruskal–Wallis Pairwise Comparisons for Intrinsic Motivation

Through pairwise statistical analysis, it was determined that participants in the control condition ranked the motivating factor of MiniNauts significantly lower than that of the no-feedback, and feedback conditions. It can therefore be stated that participants exposed to the gamified version of reading comprehension activities felt that the game motivated them more than participants not part of the gamified condition.

6.4 Game Experience Questionnaire Results

As previously mentioned in Section 3.4.2.2, a Game Experience Questionnaire (GEQ) was provided to students at two intervals during the course of this study. The purpose of the GEQ is to descriptively map player experiences of a system via a calibrated series of tests (IJsselsteijn et al. 2013). To ensure expediency of the survey period for teachers and students, the in-game form of the GEQ was employed, which uses a reduced set of 14 questions, which still map to the core components of player experience. As can be seen in Table 94, and Figure 6-16 the three Test Groups experienced several interesting relationships between component scores.

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	**Immersion Wk.4**	2.944	54	1.152	0.157
	Immersion Wk.8	2.556	54	1.058	0.144
Pair 2	**Competence Wk.4**	3.296	54	1.114	0.152
	Competence Wk.8	4.028	54	0.968	0.132
Pair 3	**Negative Wk.4**	2.870	54	1.186	0.161
	Negative Wk.8	3.130	54	0.896	0.122
Pair 4	**Flow Wk.4**	2.732	54	1.379	0.188
	Flow Wk.8	2.593	54	0.753	0.102
Pair 5	**Tension Wk.4**	2.222	54	1.239	0.169
	Tension Wk.8	2.482	54	0.911	0.124
Pair 6	**Positive Wk.4**	3.009	54	1.088	0.148
	Positive Wk.8	3.537	54	0.966	0.131
Pair 7	**Challenge Wk.4**	3.657	54	0.985	0.134
	Challenge Wk.8	3.176	54	1.078	0.147

Table 93 GEQ Descriptive Statistics

For the Week 4 survey, players experienced Immersion, Flow and Tension of (mean = 2.944, SD = 1.152), (mean = 2.731, SD = 1.379), and (mean = 2.222, SD = 1.239) respectively. With the components of Competence and Challenge, players experienced (mean = 3.296, SD = 1.114), and (mean = 3.657, SD = 0.985) respectively. Finally, Positive and Negative traits were described as (mean = 3.009, SD = 1.088), and (mean = 2.870, SD = 1.186) respectively shown in Table 93.

Factor	TG1	TG2	TG3
Immersion	3.083	2.833	2.917
Competence	3.278	3.194	3.417
Negative	3.3306	2.722	2.583
Flow	2.583	2.722	2.889
Tension	1.778	2.472	2.417
Positive	2.889	3.028	3.111
Challenge	3.75	3.667	3.556

Table 94 GEQ Week 4 Scoring

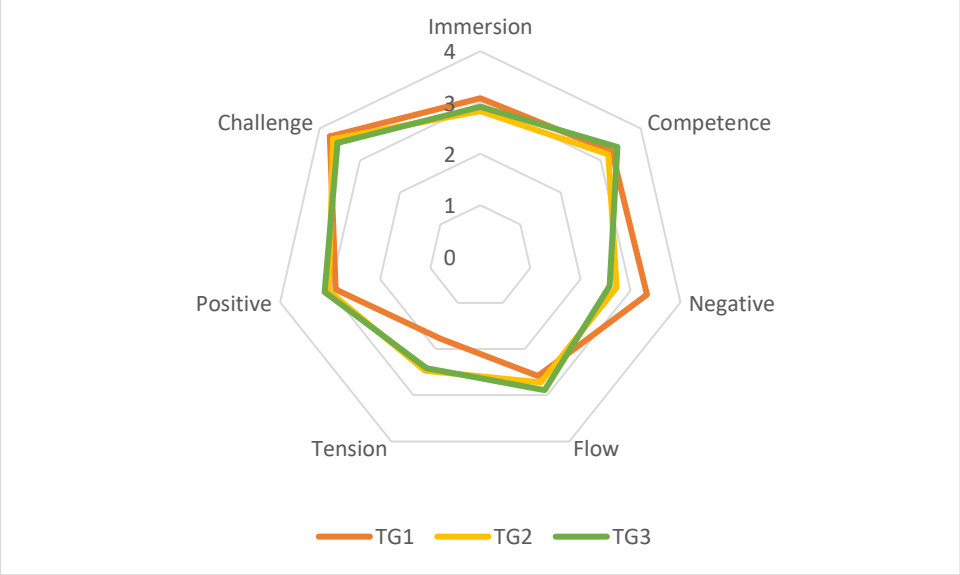


Figure 6-16 GEQ Trait Comparison Week 4

A repeat of the GEQ survey was run at the conclusion of the study, during the final play session with the system, in Week 8. Participants were again asked a series of 17 questions, rating their agreement on a five-point Likert scale, which was then summarised into test conditions. Players experienced Immersion, Flow and Tension of (mean = 2.556, SD =1.058), (mean = 2.593, SD = 0.753), and (mean = 2.482, SD = 0.911) respectively. With the components of Competence and Challenge, players experienced (mean = 4.028, SD =0.968), and (mean = 3.176, SD = 1.078) respectively. Finally, Positive and Negative traits were described as (mean = 3.537, SD = 0.966), and (mean = 3.13, SD = 0.896) details of these scores are displayed in Table 93 and Table 95.

	TG1	TG2	TG3
Immersion	2.556	2.722	2.389
Competence	4.000	3.8333	4.250
Negative	3.306	2.556	3.528
Flow	2.611	2.750	2.417
Tension	2.222	2.306	2.917
Positive	3.722	3.389	3.500
Challenge	3.194	3.500	2.833

Table 95 GEQ Week 8 Scoring

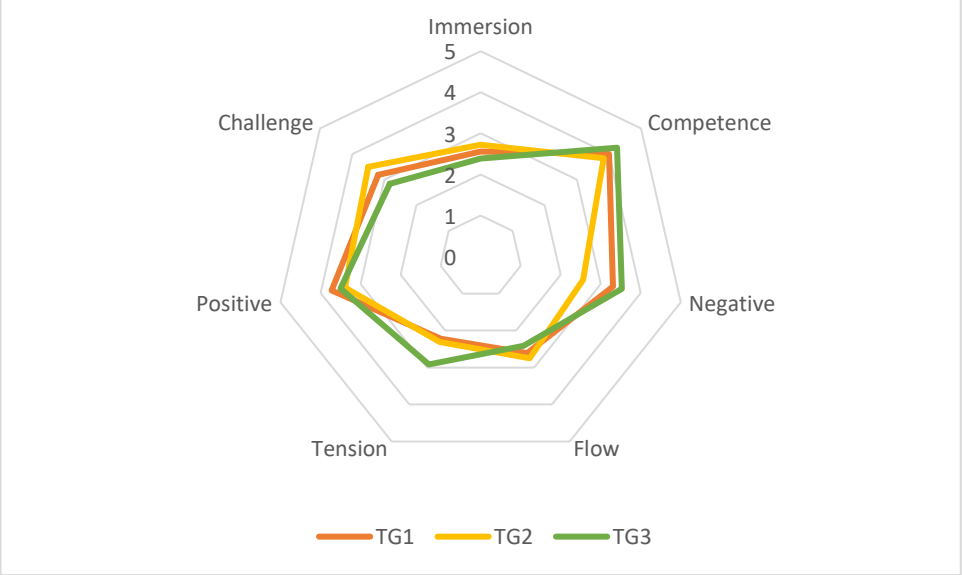


Figure 6-17 GEQ Trait Comparison Week 8

When comparing mean differences between Week 4 and Week 8 tests, several changes in mean rankings of components were measured, listed in Table 96. For the components relating to Immersion in the game, a net negative change occurred of −13.180%. A further reduction in Challenge was noted, with a net reduction of −13.177% between weeks 4 and 8. The remaining five factors of Competence, Negative, Flow, Tension, and Positive received a net increase of 22.210%, 8.718%, 5.053%, 11.701%, and 17.550% respectively for the periods of Week 4 to Week 8.

Components	Week 4	Week 8	Percent Change
Immersion	2.944	2.556	−13.180%
Competence	3.296	4.028	22.210%
Negative	2.879	3.13	8.718%
Flow	2.731	2.593	5.053%
Tension	2.222	2.482	11.701%
Positive	3.009	3.537	17.550%
Challenge	3.658	3.176	−13.177%

Table 96 GEQ Score Changes

To determine whether significant changes occurred between GEQ components for the Week 4 and Week 8 testing periods, statistical analysis was performed. Through testing for normality, it was determined that 12 out of 14 datasets violated the assumption of normality, with respect to Shapiro—Wilk testing ($p < 0.05$), as shown in Table 97. Tests which did not violate the assumption of normality were Positive Week 4 ($p = 0.058$) and Challenge Week 8 ($p = 0.056$). Due to samples being paired, and no pair of samples sharing normally distributed data, it was determined that a Wilcoxon Signed-Rank Test would be appropriate for the data analysis.

	Kolmogorov—Smirnov			Shapiro—Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Immersion Week 4	.169	54	.001	.940	54	.009
Immersion Week 8	.134	54	.018	.919	54	.001
Competence Week 4	.154	54	.003	.934	54	.005
Competence Week 8	.194	54	.000	.860	54	.000
Negative Week 4	.178	54	.000	.931	54	.004
Negative Week 8	.197	54	.000	.946	54	.016
Flow Week 4	.110	54	.099	.902	54	.000
Flow Week 8	.179	54	.000	.947	54	.018
Tension Week 4	.220	54	.000	.847	54	.000
Tension Week 8	.159	54	.002	.947	54	.019
Positive Week 4	.115	54	.075	.958	54	.058
Positive Week 8	.258	54	.000	.879	54	.000
Challenge Week 4	.155	54	.003	.932	54	.004
Challenge Week 8	.155	54	.002	.958	54	.056

Table 97 GEQ Test of Normality

A Wilcoxon Signed-Rank Test was performed on the GEQ data, categorized into seven core components, paired across testing performed in Weeks 4 and 8 of the intervention; shown in Table 98.

	Immersion	Competence	Negative	Flow	Tension	Positive	Challenge
Z	−1.933[b]	−3.074[c]	−1.207[c]	−.656[b]	−1.343[c]	−2.664[c]	−2.048[b]
Asymp. Sig. (2-tail)	.053	.002	.228	.512	.179	.008	.041

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

Table 98 GEQ Wilcoxon Signed-Rank Test for GEQ Traits

For the factors of Negative (p .228), Flow (.512), and Tension (.179), no significant variance was detected. For the factors of Competence, Positive, and challenge, significant variance was detected ($p < 0.05$); Immersion was identified as approaching significance ($p = 0.053$). Due to the lack of normality present within the GEQ, across 12 of the 14 sets, further pairwise testing is not possible between test groups for component pairs. To answer SRQ3, it can be stated that motivation did not factor significantly into the MGS and was not considered a factor in reading comprehension skill acquisition.

6.5 Post-Study Teacher Interview

A post-study interview was performed following the conclusion of the intervention to survey the opinions and feedback of the teacher participants in this study. Thematic analysis of the two teacher interviews was performed, and themes and codes are presented in Table 99.

Themes	**Student attention and enjoyment**	**Content Delivery**	**Effectiveness of tool**
Codes	Students liked the game	Simple	Performance was largest in poor performing students
	Boys liked the game	Lots of data	Everyone seemed to get faster at game
	Used the system at home	Takes time to learn	Enough feedback
	Takes time to learn	More than enough exposure time	Students noticed difference between cohort games
	Preference over in class activity		

Table 99 Thematic analysis themes and codes

The following section presents a summary of the thematic analysis identified themes, with respect to the teacher interviews.

1. **Did you find that the Mini-Game System (MGS) was overall an effective tool in your classroom?**

Teachers found the MGS to be an effective tool for use in their classrooms. Teachers identified several students who reacted very well to the system, and engaged with content after class, and at home. Teachers felt that the MGS provided an opportunity for players to test their skills with respect to reading comprehension in a novel manner. The novelty of this system in the opinions of the teachers was a significant factor in the uptake of the MGS; teachers did not feel that interest would wane significantly for the system, as several other serious games used in their classroom were popular all year long, particularly *Mathletics*.

Teachers felt that males were more interested in the system than females but did not identify a serious disparity between the two genders with respect to their overall score or general interest in the context of class-time.

2. **Do you believe that the students who used the MGS received benefits to their reading comprehension skillset?**

Anecdotally, teachers believed that their students significantly improved their reading comprehension skills over the course of the intervention. Teachers felt that students were better in class at contextualising literacy objects, inferring subtle meaning from texts, and general reading skills. Teachers believed that this benefit was mostly identified in the poorest readers of the class, with many students in the middle or upper end not changing significantly with respect to their reading abilities.

3. **In your opinion, did students personally enjoy using the MGS overall?**

Teachers believed that there was enthusiasm for the MGS across the course of the intervention. Teachers felt that students relished the opportunity to engage with non-traditional activities in the computer lab and felt that students were more motivated to engage with literacy content, if it was presented in the context of the MGS. Teachers felt that students were attached to the game store, and customisation of the character; this customisation allowed students to feel ownership over their progress, and the desire to unlock a new hat was considered an extrinsic motivator to the reading comprehension activities.

4. **Did the students enjoy the gameplay experience?**

Teachers felt that students who were previously behind in reading comprehension skills, and students who had a keen interest in video games benefited the most from the system. Teachers felt that the short nature of the mini-games fit the model of timing class activities well. Teachers could chain together as many activities as they liked, and students were able to complete them quickly and atomically. Students were not engaged in content for too long, and as such did not lose interest too fast, or get caught on complicated steps. Teachers were highly enthusiastic about the mini-game nature of the MGS.

5. **How effective do you believe the exposure time was for students using this system?**

Teachers felt the desired length of exposure of three times per week worked very well. Teachers regarded this amount of time, totalling around one hour per week on average as an appropriate length of time for this form of literacy education, as well as time devoted to reading comprehension activities in general. Teachers believed that more exposure probably wouldn't have helped many students more than it did and believed that even fewer minutes spent playing the game per week would have worked just as well.

6. **If you could perform this intervention again, what aspects would you change, to gain a greater benefit from your students?**

Teachers felt that several aspects of the research could be improved. Teachers identified the lack of long-term game content to be a potential hazard as to administering the MGS over the long-term. Teachers felt that added mini-games focusing on a broader range of skills in the literacy domain, such as vocabulary and grammar would fit the MGS well and should be included as an option in future iterations. Teachers also felt that there was a minor learning curve to using the system, particularly the Teacher Portal, and felt that some of the processes could be streamlined, particularly editing large pieces of text, and re-organising multiple-choice quizzes.

6.6 Summary

This Chapter presented data collected during the course of the previously described intervention. Through variance, and correlational analysis, several key findings were

determined, which act to answer the defined research questions SRQ1, SRQ2, and SRQ3. Firstly, it was determined that no significant differences were observed between all three cohorts, when comparing mean pre-test scores. It was also determined through statistical analysis of variance that mean variances of significance existed between TG1 and TG2, and TG1 and TG3. These variances represented significant deviations in mean post-test scores, when comparing in a between-group analysis of variance. It was determined through statistical analysis that no significant difference in mean post-test scores existed between TG2 and TG3.

Through analysis of self-reported facets of motivation, and enjoyment, several key findings were discovered. Firstly, it was determined that gender and age did not act as covariates to the game conditions. It was also determined that the factors of interest in video games, enjoyment of reading as an activity, and hours spent playing video games, did not significantly act as covariates upon the game conditions. Factors of the GEQ presented several net decreases over time (Immersion, Challenge), and the remaining components with varying levels of increasing net changes from Week 4 to Week 8. These findings will be further discussed in the following Chapter, with reference to the current body of knowledge.

7 Discussion

This Chapter will discuss the findings presented in the preceding Results Chapter, with reference to current literature. To restate the original research question (RQ): *To what extent can reading comprehension skills in students be strengthened by employing formative feedback within a system of structured mini-game experiences, and can this system satisfy the needs of students and teachers.* The purpose of this study was to present a comprehensive answer to this question, along with its constituent Sub-Research Questions, which were presented in Chapter 3, as follows:

Sub-Research Question 1 (SRQ1): *In what ways do playing a system of mini-games increase reading comprehension performance when compared to traditional presentation of these activities?*

Sub-Research Question 2 (SRQ2): *In what ways do playing a system of mini-games, with variable periods of inter-game feedback, increase reading comprehension skills?*

Sub-Research Question 3 (SRQ3): *What is the relationship between motivation, game experience, and reading comprehension skill acquisition, when reading comprehension activities are delivered via mini-games?*

Sub-Research Question 4 (SRQ4): *How can educational mini-games be delivered such that they satisfy the needs of students, teachers, and game designers?*

The following sections of this Chapter will present a discussion whereby these questions are answered.

7.1 Gamified Learning Improvements

7.1.1 *Within Group Changes*

As presented in Section 5.1 SRQ1 is informed by a pre- and post-test design, whereby participant reading comprehension skills were tested before and after an eight-week intervention. The results of these two tests were formalized into an overall measure of performance, on a scale of 0 to 100%, across the 18 questions.

7.1.1.1 Pre-and Post-Intervention

It was determined through analysis of descriptive statistics for the tests, that participants in TG1, TG2, and TG3 all presented neat improvements of mean test scores between the pre- and post-test collection periods. The statistical analysis used to determine these within-group comparisons was the paired-sample t-test. The test looks for whether a cohort of individuals achieves a mean difference in their individual mean score, against the group mean score (Trochim 2006) The null hypothesis of this test can be stated as the following:

H_0 There is not significant difference between pre- and post-test scores for participants within each of the three cohorts.

Following the results of the analysis performed in the preceding Chapter, the null hypothesis is rejected. Participants in TG1, TG2, and TG3 all experienced significant changes in pre-and post-test scores. It can therefore be stated that all cohorts gained mean score improvements. TG1 pre-test and post-test means were (24.378 ± 10.102) to (50.915 ± 22.886) respectively. TG2 saw an increase of (24.383 ± 9.163) to (69.450 ± 17.491) for pre-and post-test conditions respectively. Finally, TG3 saw an increase of mean pre- and post-test scores of (28.71 ± 11.467) to (74.07 ± 16.597) respectively. To generalize the significance of these changes, it can be stated that for an average participant, a significant change in pre- and post-test scores was observed following the eight-week intervention. Interestingly too, all cohorts saw an improvement to score where mean scores moved from > 50% incorrect answers to < 50% increase answers. It can therefore be stated that cohorts who 'failed' their reading comprehension test in the pre-test, 'passed' the test on average, eight weeks later. The result of this test indicates the general success of MiniNauts for increasing reading comprehension; in this manner, SRQ1 can be answered. All groups experienced a net increase in test score from week 1, to week 8. In this manner, it can be stated that a serious game for reading comprehension can enable greater reading comprehension scores in an eight-week period.

7.1.2 *Between Group Changes*

While the preceding section dealt with issues of pre-and post-test changes to mean scores, a second and important comparison can be made between all three cohorts. The statistical comparison, a between-group comparison, aims to determine whether variances exist between mean scores of participants, between cohorts. This test, as presented in this study was performed twice; once for the pre-test, and once for the post-test.

7.1.2.1 Pre-Test Between-Group Analysis

A One-Way ANOVA was selected as the relevant statistical tool to determine whether three or more conditions had variance lying between each condition combination. As data was considered normally distributed, as tested by the Shapiro–Wilk Test of Normality ($p > 0.05$) and confirmed as homogenous in variance by the Levene Statistic ($p > 0.05$), the suitability of a parametric test of variance such as ANOVA was deemed appropriate. For the pre-test, it was determined through statistical analysis that there existed no mean variance between all three cohorts. This result implies two interesting facets. Firstly, it can be stated that random assignment of participants to the three cohorts was considered successful, as any meaningful assignment to cohorts of participants of greater or lesser reading comprehension ability would produce unequal variances of pre-test scores, as well as changes to normality measures. Secondly, it can be stated that the pre-test did not significantly affect participants in statistically different manners with respect to pre-test scores. More simply, it can be stated that the pre-test did not act as an independent variable upon the participants. The outcome of this test demonstrates that no statistically significant covariate existed at the time of pre-testing, which is a requirement for accurate pre/post-tests, according to Dimitrov & Rumrill Jr (2003).

7.1.2.2 Post-Test Between-Group Analysis

The second main between-group comparison for the three cohorts was the post-test analysis of variance between mean scores. Through the Levene Statistic ($p > 0.05$) and Shapiro–Wilk Test of Normality ($p > 0.05$), it was determined that data was considered normally distributed amongst cohorts and did not present any statistically

significant differences of variance between cohorts. Therefore, it was determined the One-Way ANOVA was deemed appropriate for the second between-group test of scores, upon post-test mean scores. It was determined through statistical analysis that a statistically significant difference in variances existed between the three cohorts. Significance of ($p = 0.02$) indicates that between groups, a significant difference in variance lies in one or more pairs. Therefore, it was deemed appropriate that post-hoc testing provide clarification to where the difference in variances lies. The Bonferroni Correction were selected to determine the location of these variances in a pairwise, combinatorial manner. Bonferroni Corrections provide an adjustment to the increasing Type I error, the false rejection of the null hypothesis, of multiple pairwise comparisons of variances (Mathworld 2017). Post-hoc testing revealed that two significant deviations in mean scores existed between the following cohort pairs: TG1 to TG2, and TG1 to TG3. The significance ranks for each were ($p = 0.017$) and ($p = 0.002$). An effect size, and power level of 0.41 and 0.67, and 0.5 and 0.83 respectively were identified. The final pairwise comparison of TG2 to TG3 did not present any statistically recognizable difference in mean score ($p > 0.05$).

The completion of statistical tests to determine if variances lay between cohorts, completes the second component of SRQ1 and SRQ2. Three null hypotheses are presented, which assert the following

$H_{0\text{-}1}$ There exists no statistical difference between TG1, and TG2 with respect to mean post-test scores

$H_{0\text{-}2}$ There exists no statistical difference between TG1, and TG3 with respect to mean post-test scores

$H_{0\text{-}3}$ There exists no statistical difference between TG2 or TG3 with respect to mean post-test scores

The three null hypotheses presented assert that changes to post-test scores will not be statistically significant between all three combinations of pairs of cohorts. As presented in the preceding section, it was found that a statistically significant difference lay between cohorts 1 and 2, and cohorts 1 and 3. Therefore, we can reject the null

hypothesis $H_{0\text{-}1}$, and state that a statistically significant difference exists between gamified conditions, and the non-gamified condition. The interpretation of this result presents several interesting factors. With changes to the mean scores of TG1 and TG2, it can be asserted that participants in the gamified cohort experienced a significantly higher mean post-test score than those of the non-gamified condition of TG1. As mentioned previously, mean scores for TG1 were (50.917 ± 22.886), and (69.450 ± 17.491) for TG2. Participants in the gamified condition had significantly higher post-test scores on average, than those in the non-gamified condition. It can be stated that the intervention, may have influenced an increase in student reading comprehension scores, with medium effect sizes, according to (Cohen 1980), and moderate power levels of 0.76 and 0.83 respectively. This corresponds with extant literature claiming that higher levels of integration and coupling of learning mechanics to game mechanics can provide greater opportunities for learning, as was discussed in detail, in Chapter 2.

Secondly, a statistically significant variance between mean scores for the non-gamified TG1 cohort and gamified with feedback TG3 cohort was identified. Mean post-test scores for TG3 were (74.072 ±16.597). It can be stated that a statistically significant difference in mean scores lay between the two cohorts, with the gamified cohort of TG3 having a significantly higher mean test score, when compared to TG1. With respect to causality, it can be stated that the intervention period may have significantly influenced the outcomes of the reading comprehension tests, amongst the two different conditions. We can therefore reject the null hypothesis $H_{0\text{-}2}$. As identified with the TG1-TG2 comparison, this result vindicates the thinking that serious games coupling learning mechanics with game mechanics tightly, can produce greater outcomes than disparate learning and game mechanics. As both tests produced a statistically significant result, it can be stated that gameful conditions when compared to the control exhibited higher level of test score change; indicating the independent variable of gamefulness may have been a factor in this change. Factors such as feedback mechanisms between TG2 and TG3 will now be discussed.

The final pairwise combination of cohorts presented in the Bonferroni post-hoc test was that of TG2 and TG3, whereby the two-gamified condition were tested for significant variances in post-test scores. Mean post-test-scores for TG2 and TG3 were

(69.450 ± 17.491) and (74.072 ±16.597) respectively. The null hypothesis H_3 states that no statistically significant difference between TG2 and TG3 post-test scores exists. Statistical tests on variance yielded a non-significant result of ($p > 0.05$), and we can therefore accept the null hypothesis $H_{0\text{-}3}$ and state that the intervention period did not influence outcome of post-test scores. To interpret this result, several factors must be considered. It can therefore be stated that enough data exists to answer SRQ2: namely, there is no significant effect between inter-game feedback and reading comprehension test scores. The ramifications for this finding will now be discussed.

Firstly, recall that TG2 and TG3 both presented the reading comprehension material to participants in identical manners. The difference between cohorts lay in the presentation of feedback mechanisms within the reading comprehension activities. It can therefore be determined that the independent variable tested between the two conditions of TG2 and TG3, did not generate a significant difference to post-test scores. This result runs contrary to extant literature on feedback mechanisms, and their usefulness for generating changes in learning outcomes. As identified in Section 2.2.1.4, feedback has been successfully used in a variety of serious games applications to produce net increases in learning outcomes. The result of this testing therefore produces several possibilities as to why feedback mechanisms in MiniNauts did not produce greater learning outcomes which were expected from the literature. Firstly, the effect size of this test condition was low ($f = .27$) along with a lower power level (.35). This may be remediated by larger sample sizes in the future, as the size of the test groups may have been a factor. However, the remaining test conditions did produce a positive result, with a moderate effect size and power. It is therefore believed by the researcher, that feedback design may be a contributing element of this result. With respect to design, feedback mechanics in MiniNauts were entirely automated. Extant research into feedback as a mechanism for learning outcomes typically describes feedback as a manual process for the learner, between themselves and the instructor (Zebel et al. 2013; Hattie & Timperley 2007). MiniNauts, as required by factors such as time, automation, and teacher preference opted for an automated system of providing feedback to the player. Recall that MiniNauts provided opportunities for feedback presentation in TG3 of scores, personal bests, long term achievement, and individual activity feedback.

Students were not presented with customised feedback, and as such, may have not received the required level of feedback needed for significant learning outcome changes. Self-reported attitudes towards feedback will be further discussed in Section 7.3.1. The level of feedback required to meet this need is however unknown and would need to be addressed in future research. One possible approach would be to more tightly integrate feedback into the serious game mechanics, potentially in real time, at the game-feedback loop level, to provide players with instant feedback, and not just after a (relatively short) play session. The rate at which feedback is presented to players, along with the granularity, and tailoring to the individual, to elicit significant changes in learning outcomes is not yet known. MiniNauts provided feedback opportunities to players at an average rate of once every 2 minutes 27 seconds. It would appear therefore that whilst feedback opportunities were regularly presented to the user, as recommended by Kluger & DeNisi (1996; 2007), they did not appear to affect significant changes. It can therefore tentatively be stated that the rate at which feedback is presented to the player is only one factor of the feedback mechanism for serious games. This corresponds to research by Zebel et al. (2013) that states that meaningful feedback is just as important as the regularity at which it is presented; a more reflective and contemplative period of feedback presentation could also provide students the space to better contextualise and re-integrate the feedback into their personal mental models. Future research may investigate more meaningful forms of feedback for the player, however the balancing act of the meaningfulness of feedback, the regularity that it is produced and presented to the player, and the time constraints of the teacher as the gatekeeper to this feedback must be explored in further detail in future research.

7.2 Demographic and Enjoyment Factors

As MiniNauts is considered a serious game, the factors of enjoyment and motivation compose one of two integral components. While educational or training potential are key factors in determining *what* a student may learn or practice, the manner in which *how* they learn with respect to entertainment is a key factor (Cowley et al. 2013)

7.2.1 *Gender and Age*

To determine whether certain demographic factors such as age or gender influenced changes in post-test scores, or the initial achievement of pre-test scores between and within cohorts, analysis of variance was performed.

7.2.1.1 Gender

For the demographic factor of gender, the three cohorts had a male/female split of 30 and 24 respectively. Through analysing the Shapiro–Wilk test of normality, and the Levene Test of Homogeneity, it was determined that gender was normally and homogenously distributed. This result indicates that participants were well categorized into the three cohorts, and ensured randomisation, as required by pre/post-test empirical testing(Dimitrov & Rumrill Jr 2003).

7.2.1.1.1 Pre-Test Gender Independent Samples T-Test

An independent samples t-test was performed on the three cohorts, with gender as the listed factor, to determine if pre-test scores varied significantly between male and female participants. It was determined that as ($p = 0.277$) no significant differences lay between pre-test scores for males and females. This result indicates that participants classified into the cohorts did not suffer from any gender bias with respect to scores. While gender has been identified as a significant area for research with respect to the disparity between male and female school performance (Weis et al. 2013) this research did not indicate that participants in either listed gender were more likely to score higher or lower on pre-test reading comprehension tests. This factor is highly specific to the individual classrooms of the students and is quite possibly a confounding limitation of the cohorts selected for this study. This does not illustrate a flaw in the methodology but could be not entirely representative of the wider population of Tasmanian students.

7.2.1.1.2 Post-Test Gender Independent-Samples T-Test

An independent samples t-test was performed between the three cohorts to determine if significant variances in mean post-test scores was observed, when factored by gender. Scores where considered homogenous and normally distributed, as tested by the Levene Test for Homogeneity, and the Shapiro–Wilk test of normality. A

significance value of (p = 0.779) was observed, which indicates no significant variance exists between males and females with respect to gender. This result, while indicating that males and females did not experience differing levels of improvement with respect to post-test score, the result matches some observations by similar studies (Papastergiou 2009; Lucas & Sherry 2004; Chou & Tsai 2007), most notably that no discernible difference in post-test scores appears between genders when administering serious game testing. Whilst the extant literature describes a significant problem with young males and reading comprehension scores (Henry et al. 2012), net changes for both male and female students was observed, and no significant variance in these scores was identified at the time of post-testing. This can lead to a partial answer for SQR4, whereby the needs of students, that is all students, can be satisfied by this system. On a gender basis, the system did not affect the genders in significant ways, such that it can be asserted that the MGS does satisfy the needs of both genders. Whilst young males in this study did not appear to suffer from a gender bias with respect to literacy, there exists the possibility that the sample size did not accurately Chapter a wide-enough cross section of the Tasmanian primary school population. Furthermore, the selection of the target school for this intervention was mediated by factors of convenience and availability for the researchers, and school staff; future studies may include a wider sample size across a broader range of socio-economically ranked primary schools.

7.2.2 *Time Spent Playing Video Games Recreationally*

Factors such as a pre-existing inclination to video games have been identified as motivating factors for students to be more interested in serious games (Landers & Callan 2011). Time spent playing video games recreationally has been identified as a factor by which interest in video games, may be correlated to increased motivation (Connolly 2012a). Correlational analysis was deemed necessary to determine if interaction with video games recreationally is tied to an increase or decrease in performance in post-test scoring of the video game. The relationship between time spent playing video games recreationally against post-test score was visually inspected. A violation of the linear

and monotonic relationships between these two factors was identified, and as such, the assumptions for valid correlational analysis were not met.

While a relationship between time spent playing video games recreationally, and post-test scores was not identified, the possibility of variances in post-test scores factored against time spent playing video games remains. To this end, a One-Way ANOVA was performed to test the previously mentioned consideration. Significance was identified as ($p = 0.770$), indicating that no significant difference between mean post-test scores was identified when factored against time spent playing games recreationally. The result of these two tests conforms to several assertions in the literature; while interest in video games has been identified as a factor in motivation to play serious games, few studies have tied this factor directly to an increase in achievement at educational video games (Garris & Driskell 2002). This research confirms the assumption that interest in video games may not present a covariate to the learning outcomes of test scores in a reading comprehension intervention.

7.2.3 *Age*

7.2.3.1.1 Age Pre-Test

To determine whether significant variances lay between participants when factored by age, regarding pre-test scores, statistical analysis was deemed necessary. A One-Way ANOVA was performed to ascertain any variance. Normality was assured, with a Shapiro–Wilk significance of ($p > 0.05$). Homogeneity fell below the threshold, using the Levene Statistic, but due to its proximity to the significance limit ($p = 0.044$), the test proceeded in any case. A significance value of ($p = 0.177$) was generated for the data used in this test, and as such, the assertion can be made that age did not factor as a condition for significance variance in pre-test scores.

To interpret this result, it is important to understand the context in which the data was gathered. As previously described, participants in this study were composed of students in two year 5/6 classes. Mean age for participants was 11.63, SD= 0.59, with a minima of 10, and maxima of 12. During these classes, educational content is not routinely distributed by age, or subsequently grade level. The lack of significance for participant post-test scores when factored by age indicates that the stratification of

educational instruction, practice, and subsequent recall of comprehension skills is not of measurable significance. The pre-test therefore acts to validate the grouping of participants into cohorts, regardless of age.

7.2.3.1.2 Age Post-Test

To complement the validation of age not influencing initial pre-test scores, post-test factorization by age was performed, using a One-Way ANOVA. Normality and homogeneity were considered within acceptable limits for use as input for a parametric variance test. Through testing, a significance level of ($p = 0.209$) was achieved, indicating that age was not significantly variant in the population, when observing post-test scores.

The lack of significant variance between participants within the population, when stratified by age, with respect to post-test scoring indicates several things. Firstly, it can be asserted that age did not influence how well students performed at the game, and the subsequent post-test. As performance increase significantly for all cohorts between the pre- and post-test conditions, it can therefore be deduced that age was not a significant contributing factor in this change in reading comprehension performance. Secondly, it can be asserted that no certain age groups increased in performance at a relatively faster or slower pace than other groups. As age was factored at the integer level i.e. 10, 11, and 12, the categorization of participants into three cohorts did not present a significant mean difference in post-test score. SRQ4 can be further answered to, whereby MiniNauts was age-agnostic in that the mechanics of the game did not exclude those of different age ranges; the teacher was in control of content delivery, so the burden of accessibility based on age largely fell on them. Tentatively, it can be asserted that MiniNauts satisfied the needs of an age distributed cohort of at least three years. Future research using the MiniNauts serious game, or the methodology to generate such a game, may choose to include cohorts of a wider variation in age. MiniNauts, as a content agnostic platform can provide learning content to any age rage within reason; future studies may take this factor into account, and test whether age may act as a covariate on test core changes.

7.2.4 *Interest in Video Games and Other Factors*

While time spent playing video games may indicate an increase affinity with video games, consider the case of a casual gamer, who may only play video games for one hour per week, but still regards them as enjoyable and interesting. For this purpose, correlational analysis between the factors of self-reported interest in video games, and post-test scores was identified as a possible area of interest. Reported interest in video games was collected via the pot-intervention survey, with participant responding to a five-point Likert scale of enjoyment statements. The relationship between these factors was plotted on a Q-Plot, and visual inspection indicated that no monotonic or linear relationship existed. Therefore, a violation of the principals of valid correlational analysis negated the need for further relationship testing.

A Kruskal–Wallis analysis of variance was performed to determine if variance existed between post-test scores, factored against self-reported enjoyment of video games. Through analytical test, a significance value of ($p = 0.707$) was determined, indicating that no significant variance existed between factored groups of participants with respect to interest in video games. This result corroborates the findings presented in Section 7.2.2, whereby serious game success does not appear to be affected or mediated by interest and involvement with video games external the research. Whilst serious games have been identified a strong mechanism for learning and literacy education, the pre-inclination of interest in video games does not appear to affect significant effects on test score changes. It can be stated therefore that whilst participants may enjoy and choose to prefer serious games over traditional techniques, their broader experience with video games does not appear to affect the strength at which serious game interventions can change their test scores.

Students were surveyed their use, preference and preferred genre with respect to video games, during the post-study survey. It was identified that males and females shared dissimilar interests in video game genres, with sports, and racing games being the preferred genres of choice for males, and puzzle/building games the preferred genre of females. Hours spent playing video games too was different for males when compared to females; males engaged with video games for an average of 2.17 hours per

day longer than females. Further to this, the differences continue minorly with respect to the preferred platform of choice to play video games recreationally. Males preferred consoles of most other device platforms, with the opposite of smartphones being true for females. The implications of these preferences are of interest. It can be stated that gender differences in males and females have been identified in the commercial, and educational gaming sectors (Hartmann & Klimmt 2006; Greenberg et al. 2010). Extant research discusses the competing interests between males and females with preference to video games (Hartmann & Klimmt 2006; Greenberg et al. 2010). This research illustrates several of the differences between males and females with respect to video game preference. It can be said the females in the cohort were less likely to play games outside the home and played these games on smartphones more often than males. The outcome of this finding points to the limits of MiniNauts as a platform. If females and males prefer smartphones and consoles for their gaming experiences, serious games like MiniNauts—administered through a PC—may not be the most effective platform for student interest. This fact does need to be grounded in the resources of schools, and the expectation that schools need to buy custom gaming hardware is and should be of lower priority, than using the limited resources the schools may already have.

7.2.5 *Enjoyment of Recreational Reading*

While enjoyment of video games has been identified as a possible factor for increasing motivation for users of serious games (but not test-scores, as shown in Section 7.2.2, and Section 7.2.4), a second factor has been identified by the literature as a possible factor in the achievement of improvements to literacy (Cox & Guthrie 2001). In this study, enjoyment of reading was measured on a five-point Likert scale, as part of the self-reported survey following the conclusion of the intervention. To determine whether reading enjoyment was related to increases in post-test scores for participants, correlational analysis was identified as an appropriate test. Through visual analysis of a Q-Plot, it was determined that the two factors violated the assumption of linearity and did not lie in a monotonic relationship. Therefore, it can be concluded that no linear relationship exists for participants with respect to reading enjoyment and post-test scores. Students identified on average 0.8 hours of recreational reading per day for

males, and 1.29 hours of reading per day for females. This result appears consistent with the literature, stating that females read on average longer than young males, by an average of 10 minutes (Media 2014).

While a relationship between reading enjoyment and post-test scores was not identified, the possibility of variance between cohorts with respect to the same factors was identified as a suitable area of analysis. A Kruskal–Wallis Test was performed upon the previously mentioned items of data, with cohort grouping acting as the factoring variable. It was determined that a significance value of ($p = 0.923$) did not satisfy the requirements for significance of variance. While enjoyment of reading may be considered a factor with respect to reading comprehension classroom activities, it is important to note that the two reading environments, mainly home and at school, employ very different reading texts and scaffolding in which these texts are read (Anderson et al. 1988). Importantly it must be noted that while a person may enjoy reading for pleasure, this may not preclude them to enjoy reading in a more coercive format, such as during class, and particularly upon texts they did not select themselves. Whilst anecdotally, females are considered stronger readers at earlier ages than males (Gambrell 2011), as gender was not a covariate on test scores, listed in Section 6.2.1, it can be surmised that whilst females may read more, it did not affect reading comprehension scores in the post-test.

7.3 Motivational Factors

7.3.1 *Usefulness and Enjoyability of Feedback Features*

To determine the manner in which MiniNaut's feedback features are perceived by participants, two analytical tests were performed upon the post-study survey. Through statistical analysis, it was determined that participants in the control and no-feedback condition rated the usefulness of feedback features as lower, than participants in the feedback condition. While students in all three cohorts were exposed to feedback in the form of the game loop, only students in the feedback condition were presented with a detailed breakdown of their performance across the session, and the intervention.

Therefore, participants rated these extra feedback features as useful, when compared to participants who did not have these features. It is important to note, that the survey provided students to rate the usefulness of the game providing feedback to them, and not specifying what kind of feedback it was referring to. This result acts as a validator, as the work of Abrams & Gerber (2013) assert that explicit, elaborative feedback is an integral component of game based learning. The presence of this feedback for participants formed part of a basic period of reflection for students, between individual game sessions. Participants were explained where their scores came from, and how it was positioned with regards to previous attempts at that individual task. The reflective period between sessions allowed participants to position their feedback contextually regarding their overall progress. The fact that participants in the gamified feedback group perceived the usefulness of feedback features as of greater use, than more than most of their peers, combined with the knowledge that those participants outperformed their peers, strengthens the case for answering SRQ3 in the affirmative. In the general case, feedback may have a more significant affect, in future iterations of this system, by integrating greater abstractions of feedback to the student, and perhaps be integrated in game. The presence of feedback between mini-games was to promote a period of reflection, but this may not be the only option for presentation. More immediate, in-game feedback could be used in the future. This positive result does contrast with the findings presented in Section 7.1.2.2, which found that no significant difference between TG2 and TG3 existed with respect to test scores; recall that the independent variable separating TG2 and TG3 was that of regularly spaced feedback opportunities. The outcome of both tests presented in this section, and in Section 7.1.2.2 indicate that whilst reading comprehension scores were not significantly affected by feedback mechanics, the enjoyment of these mechanics was identified as being more useful than their peers in TG3 when compared to TG2 and TG1. Measuring reflection is a difficult task, as metrics such as time spent interacting with feedback may not tell the whole story. Measures such as self-reported usefulness and enjoyment can provide a picture however of student attitudes towards feedback. The survey asking whether user's found feedback useful did not name features specifically, so it was up to the user to determine what the term 'feedback' meant in the context of the serious game. Recall that in Section

2.2.1.3 feedback features could be described as implicit gameplay feedback, or external features or milestones; in the case of MiniNauts, these took the form of personal bests, and customised feedback from the teacher for each activity. It can therefore be stated that students in TG3 identified these features as unique and rated their usefulness higher than their peers in TG2 and TG1.

A second analysis of variance was performed to determine whether the self-reported levels of user enjoyability varied significantly between cohorts. A One-Way ANOVA was performed upon user scores of feedback enjoyment, factored by cohort. The result of this test indicates that no significant variation lay between participants with respect to their perception of the enjoyability of feedback features. The results of this test can be interpreted in two main ways. Firstly, it can be argued that participants who were exposed to elaborative feedback features did not find them any more enjoyable than participants who were not exposed to these features. It can then be argued that these features, even when present in a more explicit manner, are not perceived as more or less enjoyable than feedback features that are not explicit in TG1 and TG2. Secondly, as participants of TG3 stated that the feedback features were more useful significantly when compared to their peers, the lack of enjoyability of these features means that a disconnect exists between enjoyment and usefulness with respect to explicit elaborative feedback.

7.3.2 *Feelings of Improvement*

An important consideration in the field of serious games, is the feeling of improvement that students feel in both their measurable performance, and also their attitude toward the educational activity (Garris & Driskell 2002). To this end, a Kruskal–Wallis variance test was performed to determine the extent to which variance differed between participants in the three cohorts.

Participants' sentiments of improving their reading comprehension skills can be summarized as a feeling of neutrality. While, the control condition had a lower overall feeling of improvement, with respect to no-feedback, which also was lower than the feedback condition, there was no significance in these variances. It can be therefore

simply stated that participants believed the game neither helped, nor hindered their performance at reading comprehension.

While students did not feel strongly in either direction as a mean, the result of this test indicates that participants achieved higher results overall when comparing post-test scores, but this was not noticeable for the students. As pre- and post-test scores were not distributed to the students, feelings of measurable gains in performance, statistically significant for all three cohorts, was not felt strongly. It can therefore be asserted that while students may perform better at reading comprehension activities, the ability to internally register that a change has occurred may be hard to internalize, or less likely to occur on shorter time scales. The presentation therefore of learning achievement over time as integral to short-term motivation is therefore called into question. Whilst long-term changes in motivation may require greater presentation of longitudinal measurements of performance, the results of these tests indicate that no significant effects occurred, when presentation of achievement was removed for certain test groups.

7.3.3 *Game and Reading Difficulty*

The conception of difficulty in MiniNauts with respect to difficulty, is an important consideration of this research. Participants were asked during the post-test survey to appraise the difficulty of the reading and gaming activities present throughout the game. The purpose of this appraisal was to determine whether participants in the different cohorts experienced reading or gaming difficulties of different mean levels. Two Kruskal–Wallis analysis of variance tests were performed, to determine if significant variance between mean appraisals of reading and difficulty respectively, existed between the three cohorts. Game difficulty was ranked as control (26.39 ± 27.749), no-feedback (27.78 ± 22.506), and feedback (22.22 ± 25.565). It was determined that no statistically significant variance in means existed between the three cohorts. This result also vindicates the use of maths questions for the control group, as opposed to literacy questions. Interpreting the scores of participants demonstrates that as a mean, participants ranked the game as being *quite easy*, as described on. Participant ranking of game difficulty is an important consideration with respect to serious games. According

to flow theory with respect to serious games, participants must maintain a sense of struggle with a game to be pushed to achieve (Pavlas 2010; Kiili 2005). In opposition, yet complimenting this difficulty, participants must also maintain a sense of enjoyment, balanced against the difficulty of the game. As previously described, participants maintained that MiniNauts was moderately enjoyable, across all three cohorts. It can therefore be stated that MiniNauts challenged players to achieve goals in the game, both reading and gameplay, while also allowing players to have fun simultaneously. This contributes to answering a portion of SRQ4 in the affirmative, with respect to player enjoyment, and overall motivation contributing to the success of the system. Players' needs can be met by the MiniNauts system.

Participants, as previously described in the preceding Chapter, rated MiniNauts as containing somewhat easy, or neutral difficulty, when asked to rate reading difficulty during the post-study survey. It is important to note that all reading content provided to participants during the study was generated by teacher participants. It was therefore up to teacher participants to decide upon the level of difficulty for reading activities presented to participants during the intervention. Reading difficulty was considered not of significant variance between all three cohorts, as determined by a Kruskal–Wallis variance test. This demonstrates that participants did not perceive reading difficulty differently dependent upon the independent variable of game exposure. This lends credence to the homogeneity of the random assignment of participants into cohorts, but also raises the point that participants who were helped more through the game's feedback mechanisms, did not perceive the reading to be any easier than participants who were not provided with this support. (Lieberman 2010) describes feedback mechanisms as providing opportunities for students to engage with reading at a more meaningful and continual level, and as such, potentially have lower ratings of reading difficulty. While the result presented in this study indicates that the three cohorts did not experience any significant variance with respect to reading difficulty self-assessments, two points should be noted. Firstly, as reading content was designated by teacher participants, the desire to reduce workload for lesson plans is apparent. Creating and crafting difficulty content for a video game system is time consuming, and as such, the desire to ensure that all students had the potential to interact with the system and

gain something from it, may have lowered the content difficulty. Secondly, while participants did not rate the reading content as significantly different between cohorts, it must be again stated that a significant variance in reading comprehension scores appeared between participants in the feedback and control conditions. It can therefore be argued that participants may not have judged reading content as being difficult in nature, but instead a difference was occurring with participants being cognizant of this effect. Kiili (2005) describes the state of flow and unconscious learning, as an act through which training and mental schemas can be solidified, without the participant being aware that such a change is taking place. This lends credence to the theory that participants were continually learning and improving their reading comprehension skills, without being aware that such a change was taking place, as evidenced by post-test reading comprehension scores.

7.3.4 *Game as a Substitute*

An important and foundational component upon which this research is based upon, is the assumption that video games can act as an effective substitute for traditional literacy methods in a classroom. This assumption has a weight of research behind it, (Connolly 2012b; Gee 2005; Garris et al. 2002; Backlund & Hendrix 2013), but the core assumption needs to be revisited once research touches upon an area of education that has not yet been tested via serious games research. To determine whether MiniNauts effectively acted as a substitute for students, participants were asked to rank their agreement to a statement recommending the game as a viable substitute to traditional learning materials. It is important to note here that while participants who were exposed to the intervention may or may not have ranked the game as an effective substitute, other stakeholders in student education, such as teachers and parents may have competing or dissimilar views to student participants. However, this research is primarily focused upon quantitative categorization of student performance via testing, along with fewer subjective measures. Therefore, teacher and further afield stakeholder's opinions of the game as a substitute are left to future research.

Through statistical analysis, it was determined that a significant difference in student ratings of the game as an effective substitute for traditional reading activities

existed between control and game conditions. No significance relationship was found between the two feedback conditions, thereby answering SRQ2, wherein feedback on reading performance does not present a significant difference in learning outcomes for students. Furthermore, mean rankings of game participant's feelings of MiniNauts as a substitute were higher as an average, compared to those of the control groups. This difference in means suggests two possible points of interest. Firstly, it can be stated that participants who engaged with MiniNauts, and were exposed to gamified content, felt that the game is better to use than traditional activities. This outcome lends credence to the theory that participants are more interested in activities where gaming content is more tightly integrated into the gameplay, as opposed to games that present the gameplay as a separate object within the game experience (Frazer et al. 2007b). A second point of interest with this result lies in the interaction between the game acting as a more useful substitute for reading comprehension activities, and the perceived difficulty in reading and gameplay. As previously described, participants did not rank the game as either difficult with respect to reading or gameplay amongst the three cohorts. Yet as mentioned, participants did believe that the game could act as a better substitute in different degrees amongst cohorts. It can therefore be stated that participant ideation of game and reading difficulty was not tied to how much they viewed the game as a worth substitute to traditional reading activities. It must be noted that novelty bias is an important consideration with any intervention, particularly those involving video games (Girard et al. 2013). While novelty can be considered a factor in the enjoyment factors of this study, it is important to note that if novelty were a significant driver of participant recommendations of MiniNauts as a substitute, this interaction would exist between all three cohorts, and not the two gameful cohorts. A chance exists that the gameful cohorts are more inherently novel than the control game, but the researchers believe that this is an outside chance.

7.3.5 *Student Effort*

An important consideration to the efficacy of serious games used as tools for literacy education and intervention, is the ability by which serious games can motivate participants to perform certain actions for more sustained periods, or with a greater

sense of enjoyment and purpose (Garris & Driskell 2002; Paras 2005). It is important to note that student effort in this study was reported explicitly via a subjective survey. Effort was secondarily measured via post-test achievement scores for reading comprehension as an indication of performance, but links between performances in the two tests does not accurately reflect how well students concerted their efforts within the video game.

To assess the subjective opinions of student effort, a question within the post-study survey asked students to rate their feelings of effort towards MiniNauts upon a five-point Likert scale. Through statistical analysis, it was determined that participants did not vary with their self-reported scores of efforts amongst the three cohorts. Descriptive statistics upon the dataset indicated that the mean score for all cohorts fell within the 'moderate effort' range on the Likert scale, suggesting that participants believed that their efforts were neither small or large with respect to MiniNauts. As previously noted, the gap between student performances with respect to post-test scores indicated that participants did experience differences during the intervention period that allowed students in the gameful cohorts to significantly outperform those in the control cohort. It can therefore be stated that self-reported assessments of effort do not correlate to the acquisition of greater reading comprehension scores with respect to MiniNauts. As presented in Section 2.2.1.1, the two-dimensional approach to flow describes the user as being more immersed when their skills are challenged, their efforts remaining at a constant over time; this is due to the second dimension of challenge increasing as familiarity correspondingly increases. The result of this test therefore vindicates the assertion that for a flow state to be maintained, effort needs to remain stable over time, as challenge and familiarity increase. The measure of flow will be further discussed in Section 7.3.9.

7.3.6 *Enjoyability*

As serious games aim to blend fun and enjoyability with a secondary purpose, the assessment of fun with these serious games is not to be undervalued. Enjoyability of MiniNauts was ranked via a five-point Likert scale, which attempted to capture participant's opinions concerning how well MiniNauts engaged with their sense of

enjoyment. Participants were asked to rank the game from one to five, with one being "not at all enjoyable", and five being "highly enjoyable".

Through statistical analysis, it was determined that participants did not enjoy the game significantly differently between cohorts. All three cohorts ranked the game as being "moderately enjoyable" on average. This result indicates several important facts about MiniNauts, and how the intervention performed as a study. Firstly, it must be noted that while participants did not believe that the game was highly enjoyable, participants in the two gameful cohorts performed significantly higher in reading comprehension tests than participants in the control cohort. It can therefore be theorized that enjoyment, while presenting itself as a hook for participants to engage with the game, and being more enjoyable than traditional activities, was not considered as highly enjoyable on average. Students in the control cohort did not suffer for having different educational content in the game, as no significant variance was detected. Sampling for students was provided at a rate of around three sessions per week. This sampling rate may have affected enjoyability of the game, and as such, should be considered a factor in future studies. Secondly, the interaction between enjoyment and learning acquisition, while being identified as a significant factor in research (Gee 2005), does not in this circumstance appeared to affect the reading comprehension skills acquisition of participants. Participants in all three cohorts were just as likely to enjoy this game as each other, as well as believing that their individual improvements was static, as previously described, was not reflected in real world test scores. This result again lends credence to the hypothesis that participants were indeed learning and gaining skills with respect to reading comprehension, and yet this acquisition was not clear to them as they concluded the intervention, as recorded by the post-test survey. Finally, it must be stated that a limitation to the collection of enjoyment metrics exists within this methodology, while collecting feelings of enjoyment is an important measure as to the efficacy of an intervention system, feelings of enjoyment may wane over time, particularly at the end of an eight-week intervention. Future research may continually monitor participant enjoyment over time in a longitudinal fashion, which may shed light not just in feelings of enjoyment over time, but the strength of enjoyment decay with respect to literacy serious games. This however is outside the scope of the presented research.

7.3.7 *Interest in Games with respect to Enjoyment of Video Games*

To determine the efficacy of a serious games system, it must be noted that external factors such as interest and enjoyment in video games outside of the testing environment may exist (Gee 2005). Participants may have a greater affinity or inclination toward video games than other participants, and this fact may interact with participant enjoyment of MiniNauts. To this end, as part of the post-test survey, participants were asked to rate their enjoyment of traditional or serious games on a five-point Likert scale, where one indicated a "strong dislike" of video games, and five indicated a "strong like" of video games. Mean score analysis for this question resolved that participants rated a "moderate like" of video games, across all three cohorts. Statistical analysis of a correlation between like of video games and a Likert of MiniNauts was performed. It was determined through analysis that a non-monotonic relationship existed between participant ratings of enjoyment of video games, and enjoyment of MiniNauts. It can therefore be stated that participants did not enjoy MiniNauts more, if they liked video games more. This result has an interesting implication, in that while video games were rated as being an interest for participants on average, MiniNauts was considered less enjoyable than their ranking of video games as a whole. It can therefore be stated that participants did not enjoy MiniNauts as much as a regular video game on average, but that MiniNauts was only mildly enjoyable. This result is consistent with the literature, in which many serious games, while being effective at promoting their secondary cause, do not benefit from glowing feelings of enjoyment from participants (Amory & Naicker 1999; Connolly 2012b), and can contribute to answering SRQ4, by way of demonstrating the balance of player enjoyment of the serious game, and how it is not a compelling factor with the sometimes competing interests of students, and teachers.

7.3.8 *Other Motivational Factors*

Complimenting feelings of enjoyment with respect to MiniNauts, is the concept of motivation, or more specifically, that of intrinsic motivation of the serious game. While participants are potentially rewarded in playing the serious game through their

teacher, as a form of extrinsic motivation, it is also the goal of serious games research to incentivize the player to motivate themselves through meaningful interactions with the game (Garris & Driskell 2002). Therefore, it is imperative to measure whether participants felt that the game motivated them intrinsically to play. To measure this effect, a question in the post-study survey was asked, aiming to get participants to describe their feelings of motivation deriving from MiniNauts on a five-point Likert scale. Participants rated MiniNauts as containing a low amount of motivation effect for control, a neutral amount for non-feedback, and a moderate amount of motivation for participants in the feedback cohort. Through statistical analysis, it was determined that a significant difference in participant feelings of motivation was observed between the control group and the two gameful groups. No further statistically significant difference in variance occurred between the two gameful cohorts.

This result presents several important facets which will be discussed in detail. Firstly, it was found that participants were more likely to rate MiniNauts as motivating, if the participant was playing a version of the game that integrated reading content into the game content. This significant difference did not however exist between participants in the two feedback cohorts. It can therefore be stated that participants, when exposed to gameful features were more likely to rate MiniNauts as enjoyable, over participants in the control condition. As feedback mechanisms are the only meaningful difference between the two cohorts in the gameful conditions, it can logically then be asserted that the game conditions affected how participants viewed the motivating factors of MiniNauts. Secondly, it can be stated that while participants in the two game cohorts did not have a significant difference in motivational self-reporting, there did exist a difference in means worth exploring. Participants in the feedback cohort rated higher on average feelings of motivation than those in the non-feedback condition. This difference leads the researchers to believe that motivation may be tied to feedback mechanisms, although statistically, this was proven otherwise via testing. Future research may investigate the effect of feedback on motivational factors more explicitly. Finally, it was observed that participants in the control group were more likely to rate the game as less motivating than those in the other two cohorts. This cohort also demonstrated statistically lower reading comprehension scores in the post-test, when compared to the

two game conditions. This interaction therefore leaves open the possibility that gameful factors may affect how well participants are motivated, or in the case of the control group, how little MCQs presented in traditional formats actually motivate them.

At this juncture, enough data exists to answer SRQ3, with regard to the preceding sections discussing of motivation and enjoyment. It can be stated that there exists a relationship between motivation and game experience with respect to reading comprehension test scores following an eight-week intervention. It was identified that factors such as feedback mechanisms, whilst proving not effective at affecting statistically significant improvements in reading comprehension outcomes, were deemed as useful for students. Students too regarded the serious game as a better substitute for reading comprehension activities when compared to traditional activities. No relationship between interest in video games and performance in MiniNauts was identified, indicating that motivation intrinsic in video games does not carry over significantly to serious games strongly enough to effect significant change. MiniNauts was designed as a video game, containing many game mechanics, feedback loops, and tertiary goals. It was hypothesised that skill and interest in video games could translate to performance in MiniNauts. The results concluded that there was no correlation between the two factors, and in the researcher's opinion, this was not due to the fact that MiniNauts was not considered a proper video game by students. Students performed significantly better with MiniNauts, than without, regardless of their experience or interest in video games, leading to the assertion that skill in the video game does not determine whether a student will be skilful at reading comprehension. This gives credence to the assertion that MiniNauts correctly rewards skill in the desired activity (reading), and not just those of skilful gameplay. Furthermore, factors such as feelings of improvement over time were identified as not being particularly strong with MiniNauts, even while test scores indicate that these changes were occurring for all cohorts. To summarise this, it can be said in answer to SRQ3 that students can be motivated by serious games for reading comprehension, and as such their motivation toward the game—and games in general—plays a role in their perceptions of the system, but not in their overall achievement toward learning outcomes.

To answer SRQ4, it can be stated that a serious game system for literacy can be created to satisfy the needs of students, teachers, and researchers alike. As has been presented in this Chapter, students encountered significant acquisitions of reading comprehension skills during the course of the intervention. Students maintained a strong sense of ownership over the system and stated that it was of tremendous use for their class, and the motivation of students to engage with reading comprehension activities. The tool too has been of use to the researchers with respect to gathering metrics on reading comprehension skills, whilst administering reading comprehension activities through the system. It can be stated that a system of structured mini-games satisfies the needs of students, teachers, and researchers alike. Whilst causality cannot be attributed to the significant changes in reading comprehension score for all cohorts, anecdotal, and self-reported statements relate that MiniNauts was a helpful tool for both teachers and students in class time.

7.3.9 *GEQ Components*

As mentioned in Section 6.4, two GEQ surveys were administered to students across all cohorts in weeks 4 and 8 of the study. It was observed that participants across the two surveys encountered a net reduction in Immersion and Challenge. This reduction may be as a result of familiarity of the game causing participants to understand the game, story, and action in greater detail. This understanding may have resulted in lower feelings of general immersion with the video game once players had completed a certain competency of the game. For the components of Competence and Tension, players experienced a net positive change in Competence and Tension. This result may be due to familiarity with the concepts of the game, allowing players to feel more assured and less challenged; this is mediated however by the net increase in Tension, whereby students felt an increase in frustration and irritation. It is important to note however that players experienced two potential sources of difficulty, that of the gameplay, and that of the reading content set by the teacher. While players may have increased their reading comprehension skills, the ease of play increased correspondingly, with the challenged of each activity decreasing. This leads to the conclusion that as players gained experience with the system, along with practicing their

reading comprehension skills to a significant net improvement for all cohorts— as demonstrated in Section 6.1—players felt a lessening in challenge over time. Finally, the Positive and Negative affective components had a net increase of 8.718%, and 17.550% respectively. Players of MiniNauts demonstrated a more positive affective state when talking about MiniNauts in Week 8 than compared to in Week 4. Statements such as: *I felt good*, indicate that students felt an increase in positive affective characteristics over time. Students felt negative characteristics more strongly in Week 8, when compared to Week 4 however, where statements such as: *I felt bored*, indicate that players may have grown tired of continual use of the system. Students felt a net increase in the GEQ component of Flow, with a net positive change of 5.053%. Students may have felt a slight increase in flow from Week 4 to 8 due to familiarity with the game increasing the ease with which they could interact with the system. This corroborates the finding of the increase in the Competence GEQ component whereby participants felt that whilst Competence increased, they were more inclined to remain in the flow state, this is slightly contradicted however by the reduction in Challenge that was previously presented. Whilst participants may have felt that challenge decreased over time; recall that reading difficulty and game difficulty were two sources of challenge present in the MGS. This corresponds with the findings of Section 7.3.3

It can be surmised through looking at the GEQ scores for the two questionnaires, players experienced changes in feelings of difficulty, challenge and immersion over time. These changes may point to a link between familiarity with the gaming system, and a lowering of immersion within the play experience. Students did however feel a net increase in positive sentiment to MiniNauts, with a correspond increase in their sense of frustration and boredom. It is important to note however that while play experiences may have changed, net positive improvements in pre-and post-test reading comprehension scores still took place, indicating that the system still works for its primary purpose as a tool for reading comprehension practice.

7.4 Post-Study Teacher Interview Discussion

Following the conclusion of the serious game intervention, an interview was conducted with the two teacher participants who were part of this study; opinions relating to the MGS were discussed.

Teachers felt that the serious game presented a compelling and fun opportunity for their classes to engage with reading comprehension content. Teachers felt that the students who were the furthest behind in the literacy skills benefited the most from the serious game. This anecdotal opinion aligns with research stating that students the furthest behind, typically benefit the most from serious games (Susi et al. 2007). Students were identified as being comprehensively interested in the MGS and were found to desire playing the games to completion to unlock new hats. This extrinsic motivator in the opinions of the teachers motivated students to continue with the system, even if the novelty of the game mechanics was no longer as strong as in the beginning.

Teachers also felt that participants were motivated and affected in different ways, particularly students who were behind in their literacy, along with boys being the cohort most interested in the game. While the results of Section 7.2.1.1 indicate that there was no significant variation between gender and post-test score, teachers anecdotally felt that the male students benefited from the system the most, as well as being most interested in engaging with it. This aligns with extant research on the topic of gender and serious games (Connolly 2012b; Garris & Driskell 2002; Department of Education 2011), while the quantitative results of this study do not corroborate this.

Teachers felt that the MGS could be improved through a greater variety of mini-games. Mini-games that focused on a broader range of topics including vocabulary and grammar instruction were identified as potential improvements. Upgrades to the Teacher Portal to increase the user experience of editing and organising MCQs was also identified as an area for improvement.

7.5 Summary

This Chapter discussed the analysis presented in Chapter 6 in an attempt to meaningfully answer the research question presented at the start of this document.

Factors such as enjoyment, motivation, and demographic statistics were presented, and conclusions were drawn from the presence of any interactions which were revealed. It was determined that while participants performed significantly higher on reading comprehension test across all three cohorts, it was TG2 and TG3 which demonstrated the greatest growth in reading comprehension scores. Furthermore, while the control condition did see a mean improvement in test scores, it was significantly lower than those of the two game conditions. It was determined therefore that reasonably, game elements such as integrating reading content into game mechanics may have resulted in higher test scores.

Demographic and enjoyment factors were discussed in detail, to determine in what ways external factors such as age, gender, and enjoyment of video games and reading outside of the classroom may affect the overall scores of participants in the post-test. Through statistical analysis, it was determined that demographic factors such as age and gender did not preclude higher scores amongst participants in all three cohorts. Factors such as the generalisability of the wider cohort to Tasmanian schools was discussed as an opportunity for future research. It was also determined that interest in video games, and enjoyment of recreational reading also did not act as covariates to test scores. Gendered differences between platform of choice, reading time, and game genre choice existed between genders, it is in the opinion of the researchers that whilst these factors are important for the design of a *serious* game, the factors through which this affects enjoyment and test-scores is mild at best. Students listed their interest in video games, video game platforms, and when considering gender, presented different preferences for how games are played, and what genre they are composed of. Conclusions were drawn from these results, indicating that external demographic factors did not seem to influence individual participant achievement in MiniNauts, and that this is generally supported by research in this area.

Finally, factors of motivation and use of MiniNauts was described and discussed, with several key findings being presented. Generally, it was considered that while participants did not rate MiniNauts as being a more useful game, or that participant feelings of improvement did not increase with respect to the game, student scores suggest that a deeper learning is occurring. This finding is also corroborated with scores

from the GEQ. While participants may not have viewed MiniNauts as a strong deterministic cause of their improvement, for participants who could see an improvement through feedback, this improvement was evident in changes to pre-and post-test scores, and in significantly higher post test scores for TG2 and TG3. It was concluded that while participants may be unaware of their growing skills in a certain area, the serious game may be behind this improvement. The following Chapter will present a conclusion to this research and discuss opportunities for future research.

8 Contributions, Future Work, and Conclusion

The purpose of this study was threefold. Firstly, the researcher aimed to resolve the question of whether gamified reading comprehension activities could be of significant benefit to students. Secondly, it was surmised that feedback mechanisms could be a boost to student motivation, although this was not statistically significant for reading comprehension levels. Finally, and perhaps most importantly, it was of necessity to summarise the individual and group experiences of students, to make full the promise of constructivist game design.

8.1 Contributions

This research has presented the design, and application of a unique serious game platform for reading comprehension instruction. Several important contributions to the body of knowledge surrounding serious games have been made. The following section will describe these contributions, and the ramifications for the application in the serious games design research space.

It can be stated that the most significant contribution to the body of knowledge as a result of this research is that of answering SRQ1, and SRQ2; whether a serious game for reading comprehension can work better than traditional presentation of reading comprehension content, and can this content be boosted by structured feedback opportunities. For the former, it can be answered in the affirmative, and in the latter, in the negative, within reason. A significant increase in post-test scores was identified for participants in the two gameful conditions, when compared to the non gameful conditions. This positive result updates the body of knowledge toward further understanding how reading comprehension serious games, a severely underexplored avenue of literacy research, can be positively enabled. This has been identified as the key outcome of this research and is an important stepping stone for the body of knowledge to move closer to moving reading comprehension into the 21st century. The

lack of a rigid model for presenting feedback in serious games is in the researcher's opinion, a factor in the failure of MiniNaut's feedback mechanisms. The presence of no significant effect of feedback outcomes on post-test scores contrasts to previously described literature, which presented structured feedback opportunities as vital for serious learning. Whilst serious learning did occur through the sustained use of MiniNauts, feedback did not appear to be a mediating factor in this increase in performance. Whilst the way feedback was generated, automated, and presented to the player may have played a role in the lack of difference between the two gameful conditions, it is the opinion of the researchers that feedback mechanisms are highly complex, and may interact very differently based on the learning content provided. This fact has been identified as another important outcome of this research, and the manner in which feedback has been integrated into MiniNauts, and the lack of statistical significance that arose, will provide future researchers with an insight into how to implement literacy feedback in mini-games. Reading comprehension in the case of MiniNauts did not appear to significantly benefit from feedback, although factors of self-reported usefulness amongst users listed it as a tool of modest use. The automation of feedback may have played a role in the lack of significance, however the automation of reading comprehension feedback in a serious game is the first of its kind, and future research should attempt to expand and focus on improving this feedback in future serious games. Literature suggests (Section 2.2.1.3) that varying difficulty levels, encouraging participants to remain in the 'flow' state are an important factor in learning. The results presented do not appear to represent feedback, a measure of increasing difficulty adjustment, as a significant factor in learning outcomes. Whilst difficulty was increased over time, participants did not reflect that this feedback was more beneficial to them via test score. The concept of dynamic difficulty adjustment, with respect to the feedback loop, and user satisfaction should be explored further in future research.

It was identified via this research that whilst differences between male and females may exist with respect to serious game mediated reading comprehension interventions, no significant differences were found between male and female cohorts with regards to pre-or post-testing. This contribution provides two important pieces of information. Firstly, pre-testing did not pick up on any statistically significant difference between

male and female reading comprehension scores prior to the intervention. Therefore, it can be stated that the current body of knowledge claiming that males are more likely than females to suffer from a literacy problem, may not be as simple as once thought. Whilst certain factors like choice of classroom, and effect size may influence the outcome of a type II error, it can be stated at least for the two chosen classes in this study, that year 5 and 6 students did not suffer from any gender-based disparity in reading comprehension. This result may require further research in the future, and the investigation sample size may be a component in this result. While no demographic factors appeared to be at play with regard to reading comprehension scores, or enjoyment, these factors may change with a large and wider sampling of students.

Secondly, the post-test illuminates that whilst all cohorts experienced a significant increase in reading comprehension scores across the intervention, gendered effects did not account for this significant increase. Males and females responded just as well to the reading comprehension treatment across all cohorts equally. This contribution to the body of knowledge illuminates the thought that whilst males may have an anecdotal inclination to video games, the effect of serious games on overall attainment of reading comprehension scores is not a factor.

Through observing the GEQ results for this intervention, several important contributions have been made. Firstly, it appears that participants remained in a flow state for the course of the game and intervention, as self-reported. This maintenance of flow was however contrasted with an increase in frustration, and a decrease in challenge and immersion. For reading comprehension instruction, particularly for serious games, whilst MiniNauts provided a moderately enjoyable experience for participants, learning content and game difficulty did not keep pace with the needs and wants of participants. An outcome of this study therefore would assert that reading comprehension serious games must pay attention to the difficulty level of the reading content presented, as well as the game content. MiniNauts existed as a static level of difficulty, solely relying on learning content to increase in difficulty at the discretion of the teacher. Future research may wish to modify the game difficulty dynamically to moderate the level of challenge and immersion felt by the player. Perhaps automation

The design of MiniNauts also presents a novel contribution to the body of knowledge, specifically behaviourally based serious games. The field of behaviour change is fast growing, and as yet, few serious games have attempted to ground their purpose in behavioural theory; effectively none in the literacy space. This research therefore is the first major attempt at design a serious game from the ground up, using the Behaviour Change Wheel. Also, novel, is the application of the Learning Model-Gaming Mode (LM-GM) to the Behaviour Change Techniques derived from the Behaviour Change Wheel (BCW). The positive results of this study are in the opinions of the researcher, highly related to the grounded and tested methodological base of both the BCW and LM-GM. Behaviour Change theory provided a foundational base to the later abstractions of the LM-GM, and the design of game mechanics and feature. The BCW is in the researcher's opinion, provides a systematic and scientific manner for describing core problems in student behaviour with regard to literacy; this analysis is unique to this research, and the researcher believes that non-rigorous methods for designing serious games can be made redundant with the use of this framework as a theoretical and methodological underpinning. This is an important, and new avenue of research in the field of serious games, and the implications of serious games designed from a behavioural perspective cannot be ignored. Whilst this research broke new ground in serious game design with a behavioural basis, future research must explore other game taxonomies and learning/gaming models, to help extrapolate the results of this study. A single behavioural model was used, along with that of a learning/gaming model, and whilst the results proved positive for students, this design could be made even more rigorous if cross-checked against other models or frameworks.

The final major contribution of this research is in the field of complex vs mini-games, a discourse which appears to have stagnated in recent years. This research presents a novel, modular min-game system, whereby reading comprehension content can be administered through a Teacher Portal and be automated into a serious game. Assessment too is automated, and this automation appears to be a first in the serious game space, particularly for literacy serious games. It was observed that mini-games are a sound fit for reading comprehension instruction, and the extant literature disparaging mini-games must look at the effect of mini vs complex on a case by case basis, as

evidenced by this study. The application of structured feedback between game sessions is a novel contribution to the body of knowledge and fulfils a gap in extant literature calling for its placement more regularly between game sessions. While it was found that mini-games can be used successfully for reading comprehension education, the difficulty relating to the design of these mini-games regarding the lack of extant literature into mini-games means that future work must replicate the results of this study. This replication, along with other suggestions of future work is outlined in the following section.

8.2 Future Work

As presented in the preceding Chapter, it was determined that significant changes to student learning outcomes occurred amongst all participants across the three cohorts. Whether by providing extra care to this particular aspect of student's education, or simply that computers are 'more fun' than pen and paper, is for future research to illuminate. What can be said with certainty, is that students enjoyed MiniNauts, a gamified reading comprehension tool, and benefited measurably from it. This significant increase in reading comprehension whilst promising, must be followed up by longitudinal tests, observing the long-term implications of repeated use of the tool, and how well students retain their skills as time passes. Future research may be inclined to present this tool over the course of several years and observe its effects. Future research also should examine further the role of reading comprehension instruction fatigue on teachers and determine the manner in which teachers are given the opportunity to administer content. Whilst anecdotally, teachers of this study approved of MiniNauts as a tool, the long-term use, and viability of this tool has not been tested, and future research should focus on teacher usability factors going forward.

Whilst the aspect of providing elaborative feedback for participants yielded no greater outcomes than for those not presented with feedback, its role in educational gaming should not be dismissed. Automated literacy systems such as MiniNauts provide an end-to-end activity platform for students, but as automation increases, personalisation too must decrease. It is perhaps with this personalisation of feedback, for each activity, and indeed for each participant, that elaborative feedback in this

domain is required. Further, reflective periods interacting with feedback features may be extended, providing participants with greater tools for interacting with their progress, and potentially being extended by forms of teacher feedback personalised to them. Perhaps the regularity of feedback could be increased even further, to a more real time frequency, and could be synthesised by the player at the moment of activity, not just at the conclusion. Is this personalisation at odds with automation of reading comprehension, and the time constraints of the teacher? That too is for future research to decide.

Finally, demographic factors such as age, gender, and inclination toward video games, and reading must be explored further. Whilst a self-survey was provided to participants at the end of the serious game intervention, this survey may not present a widely generalisable view of the Tasmanian or indeed Australian school population. Future research should focus on rolling-out the tool to a broad range of schools, and indeed across a wider age-range to determine the effects of MiniNauts and wider portion of students in the school system.

8.3 Conclusion

To the researcher, it is of great importance in the design and assessment of serious games that student feedback and opinion is gathered and examined; as educators and researchers, our goal is to improve learning outcomes, but the goal of the student in the chair is to simply get through the school day in the most enjoyable way possible. Video games have—and we believe with MiniNauts, are creating opportunities for fun to be injected into formalised literacy education. This research has demonstrated the unique application of the BCW with regard to serious game literacy education, and its use as a foundation for rigorous game design using the LM/GM. This is an important progression of the body of knowledge on rigorous game design, and it is hoped that the BCW is examined closely and integrated with systems in the serious game space. Participants in this study rated MiniNauts as a more favourable tool than traditional reading comprehension activities, and while it was demonstrated that GEQ scores for immersion were reduced over time, learning outcomes did not degrade, and were in fact enhanced over time. This research has demonstrated, albeit in a preliminary fashion,

that reading comprehension activities can be fully automated, and delivered in-class with minimal disruption, and positive learning outcomes. Participants received the game well, rating it is a moderately enjoyable experience overall, something that cannot be said for many edutainment and serious games in the current market. It has been demonstrated in this study that mini-games can attract the attention of students and are not a simple academic write-off that they may appear to be at first glance. With the addition of greater mechanisms of feedback, reflection, monitoring for teachers, and a longer trial period, the researchers believe that the MiniNauts MGS will only improve, and continue to generate learning outcomes for students. To broadly answer the research question presented in Section 3.1 concerning the extent of serious games strengthening reading comprehension skills, and its viability: this research can answer that the extent is significant, and indeed the system can satisfy the needs of all engaged in the education of primary school students. We consider this research a success for the field of serious games for literacy.

Appendix

I. Pre-Test Materials

Paragraph Text	Question	Answers
Thirteen-year-old Angus Paradice lives on a farm in New South Wales. In 2008, he travelled with his family on holiday to Asia. In Mongolia, he saw the famous Naadam festival horseraces. All of the jockeys were children. Angus wanted to race too, so he decided to enter the 2009 competition. After he returned to Australia, Angus trained for the long distance competition by riding 22 kilometres after school each day and by running and doing sit-ups. In 2009, Angus returned to Mongolia. Some of his Mongolian friends arranged for him to ride in a 10 kilometre race for two-year-old horses, and a 15 kilometre race for five-year-old horses. Although he had a bad fall before the races, Angus finished in the top 10 in both events. His efforts won him the 2009 Young Adventurer of the Year Award.	When he first saw the Naadam festival horseraces, Angus felt:	disappointed that the jockeys were children.
		bored because the races took so long.
		inspired to take part in the races.
		nervous about the races.
	After he returned to Australia,	home town
		exercise routine.
		racing experience.
		family background.
	Angus's Mongolian friends helped him to	take part in the Mongolian races.
		find his way around Mongolia
		train for the Mongolian races.
		plan his trip to Mongolia.
	The way this text is written leads the reader to	feel jealous of Angus.
		feel sorry for Angus.
		be amused by Angus.
		admire Angus.

	Which idea best matches this text?	There is no place like home.
		Fight for what you believe in.
		Good things come to those who wait.
		Dreams can come true if you work hard.
	Why did Angus receive the Young Adventurer of the Year Award?	He won a competition.
		He made a tough decision.
		He achieved a difficult goal.
		He did something to help others
On Saturday Patrick woke up with a little shock, knowing that this was a special day. For a moment he couldn't think exactly why, and them with an excited futter of his stomach, he remembered. At ten o'clock today he was going to find out once and for all a bout Finders Keepers. He got dressed more carefully than usual, went downstairs and turned on the TV. Quickly he switched channels. Cartoons, cartoons, advertisement, man talking,	Why is this particular Saturday a special day for Patrick?	He will be allowed to watch TV all morning.
		He will get the present he has been asking for.
		He will discover an answer he has been waiting for.
		He will be able to spend the whole day with his mother.
	Quickly he switched channels.What is the	He was impatient for the program to begin.

snow, snow.. and still nothing at all on Channel 8. 'Patrick, tune it in, darling, if you're going to watch.' Judith wandered past with the newspaper under her arm and her eyes half closed. She headed for the kitchen. Patrick turned off the TV and followed. "What's for breakfast, Mum?" "We'll see," Judith murmured vaguely, plugging in the electric kettle. She blinked sleepily at him and smiled. "You look nice darling." She said. "You're all ready. But we can't go till eight-thirty at the earliest, you know. Nothing'll be open till then." Patrick's stomach lurched. "We aren't going out are we?" He asked anxiously. She began to make the tea. "Don't say you've forgotten!" she said. "I promised you, last Saturday. Your new sneakers, remember?" "Oh-oh, but I can't go out this morning, Mum. There's something I've got to watch on TV. At ten o'clock. I've got to! My sneakers'll be all right for another week," gabbled Patrick, panic-stricken. Judith faced him, hands on hips. "Patrick," she said wearily, "it's all organised."	most likely reason that Patrick did this?	He thought the program had already started
		He wanted something to do before breakfast.
		He usually watched cartoons on Saturday morning
	What made Judith think Patrick was ready to go shopping?	He was dressed and had finished breakfast.
		He was dressed more carefully than usual
		He mentioned it to her at breakfast
		He was keen to buy new sneakers
	What is the most likely reason Judith sounded vague?	She had just woken up and was still tired.
		She was thinking about Patrick's sneakers.
		She was concentrating on fi lling the electric kettle
		She didn't want to tell Patrick they were going out
		he suddenly felt ill.

	Patrick's stomach lurched.This suggests that	he needed some breakfast.
		he had woken up too early
		he really didn't want to go out.
	Why does Patrick say, "My sneakers'll be all right for another week"?	because he doesn't need new sneakers
		because he never agreed to go shopping for new sneakers
		to persuade his mother that the shopping trip can be postponed
		to persuade his mother to go to the shops earlier than she planned
On an Arctic island long ago, a stranger is approaching a village. "Papa," I yell. "Someone is coming."Papa gathers Uncle and the other men. They come to stand beside Finn, Tuaq and me in a show of communal strength."He must be from one of the groups that have already arrived at the coast," Uncle suggests.Papa nods. He doesn't take his eyes off the approaching figure. "Get	Papa's attitude towards Hulag is	timid but kind.
		wary but polite.
		confused and fearful.
		aggressive and disrespectful
	Hulag's behaviour when he arrives can best be described as	quiet.
		confident.
		unfriendly
		aggressive

Nana," he tells Miki. If the man wants to stay, Nana will decide. She's already walking towards us, wearing her priestess cape trimmed with raven feathers and arctic fox fur."Hullo-o-o," the man calls into the wind. Papa waits until he can see the stranger's eyes. Theman is not from any villages we join with on the coast. "Good morning," Papa says cautiously. "I am Hulag," the man responds. Papa doesn't say his name. Instead he nods in Nana's direction. "This is Ananaksaq." Nana is famous throughout the icelands and Papa is reminding Hulaghow powerful our village is. "It's an honour to meet you." Hulag's eyes measure Nana up and down. He doesn't look impressed. His grin says he thinks it will be easy to charm this old womanwith an oil-stained parka and dirty face. Papa leads, but Nana decides, and she has made herfirst decision. This man must wait out in the cold.	He doesn't take his eyes off the approaching figure. "Get Nana," he tells Miki. These sentences help to	create tension.
		provide clarity
		indicate surprise.
		resolve a conflict.
	Hulag thinks that Nana	is unfriendly.
		can be argued with.
		can be manipulated.
		is strong and decisive.
	In the last sentence, Hulag is referred to as This man. The main purpose of this is to show that	he is different from Papa
		the villagers have not heard his name.
		he has gained the respect of the villagers
		he is being kept at a distance by the villagers.
	Where do the characters in the story live	on the coast
		in the tundra
		by a mountain
		It doesn't say

II. Post-Test Materials

Paragraph Text	Question	Answer
'Tiffany comes to us from Tilgong Primary,' Mrs Tarrant tells the class. 'She now lives in Mittavale and will be with us for the rest of the year. I know of course, 6H, you will make her most welcome.' This is my new classroom. I can feel everything around me; the dusty warmth of the air, the rickety table under my elbow, the lino under my shoes, my backpack on my knees. I take out my old Tilgong folder,which feels heavy with its new lined pages,and I see down next to my lunch a little red present. Dad must've put it there! For a few seconds I stop feeling scared. I don't take the present out, though, just my school stuff.	Why has Tiffany changed schools?	She has moved house
		She has run away from Tilgong
		She didn't have any friends in Tilgong
		She wanted to go to a different school
	Tiffany uses the words 'my stomach is squirming' because she feels	Happy
		Scared
		Hungry
		Excited
	'and even as I'm answering and trying to smile'. Who is the narrator in this story?	Tiffany
		Mrs Tarrant
		Mrs Henderson
		Tiffany's Father
	When Tiffany sees the surprise, she stops feeling	Warm
		Lonely
		Welcomed
		Frightened
	Why isn't Tiffany listening to Mrs Tarrant's story about another new student?	She is thinking about her old school
		She knows there are lots of new students

		She is worried about going into a new classroom
		She cannot understand what the Principal is saying
Each year the Emperor watched a kite competition among the rich people who lived in his palace. Soon it was time for all the people to gather by the steps of the palace to see the rich nobles bring out their Golden Kites. Litle Hao heard drumbeats. The nobles were coming. There they were! First came Lord North Wind. His kite was like a dragon, shining golden in the sunlight. "Aaaah!" said all the people. "That one will win the Emperor's prize." Next came Lord Noble Horse. His kite was like a golden eagle with its wings spread wide. It soared into the sky. The people cheered and clapped. Last of all came Lord Black Mountain. His kite was made like	The people say 'Aaaah!' because	The sunlight blinded their eyes
		The kites glistened above them
		The dragon kite won the Emperor's prize
		The wind was too strong for the dragon kite
	Who flew a kite shaped like a dragon?	Little Hao
		Lord North Wind
		Lord Noble Horse
		Lord Black Mountain
	The emperor is most likely to award his prize to	The kite that could fly the highest
		The kite that the people liked the best
		The kite that cost the most to make
		The kit that belonged to the richest noble

the flames from a fire and there were rich jewels in its tail. The fire kit sparkled all over the sky. The people cheered their loudest. "That one wins! That one wins!" they called.	Little Hao heard the sound of	The North Wind
		A Dragon
		Kites
		Drums
	Which word tells us that a kit flew high in the sky?	Spread
		Shining
		Soared
		Sparkled
	How often did the kit contest take place?	Every month
		Once a year
		Once every ten years
		This was the first time
One hundred years ago, the people of Germany were convinced that a horse from Berlin could count and answer tricky maths questions (including fractions!) by tapping out the number with his hoof! It wasn't until a psychologist called Oskar Pfungst studied both the horse as he answered the questions, and the people asking the questions, that the mystery was	Hans was able to give the right answers to the maths questions because he was good at	Remembering previous questions
		Noticing human behaviour
		Guessing
		Maths
	One message of this text is that	Cheats never prosper
		Practice makes perfect
		Appearances can be deceiving

solved. Pfungst noticed that Hans usually got the answer right, but if the person asking the question didn't know the answer, neither did Hans. That told him that Hans was getting a signal from the person asking the question. When a person asked a question they would tilt their head slightly without even knowing it, and Hans would start tapping. When Hans got to the right answer, the person would be so excited they'd lift their head or raise an eyebrow or smile. Hans stopped counting when he saw the change in their expression. So Clever Hans was great at reading body language, but no great shakes with maths.		You can lead a horse to water, but you can't make it drink
	The text suggests that people asking Hans the questions	Deliberately cheated and gave Hans the answers
		Gave Hans the answers without realizing it
		Were not very good at maths
		Did not trust Oskar Pfungst
	Which words best describe Oskar Pfungst	Curious and logical
		Artistic and creative
		Cunning and sneaky
		warm and sympathetic
	Why was it important for Oskar Pfungst to study Hans and the person asking the questions, together?	To test the person's maths abilities
		To make sure has wasn't frightened
		To watch how they related to each other
		To get information about Hans' background
		Emphasis how clever Hans seems to be

	The phrase "including fractions!" in the first paragraph is used to	Show that Hans' math skills were limited
		Question how Hans was able to express the answers
		Describe the difference between fractions and other maths questions
	Which piece of information lead Oskar Pfungst to solve the mystery?	Hans used his hoof to tap out the answers to the questions
		Hans had convinced the people of Germany that he could count
		Hans only knew the answer if the person asking the question did
		Hans could do fractions as well as other kinds of maths

III. Student Survey

How fun was the game?	**Not Fun – Very Fun (1–5)**
Do you like the genre (type of game) of MiniNauts?	Not at all – Yes, I do (1–5)
How much did you feel that you learned?	Nothing – A lot (1–5)
How hard was the reading part for you overall?	Easy – Very Hard (1–5)
How hard was the game part for you?	Easy – Very Hard (1–5)
How much did the game help you to remember things you learned in the game?	It didn't help at all – It helped a lot (1–5)
How much did the game did the game tell you how you were going?	It didn't tell me at all – It told me a lot (1–5)
Was seeing your score and progress helpful?	No, it wasn't helpful – Yes, I loved it (1–5)
Did you get better at the game over time?	I didn't get better – I got better a lot (1–5)
Do you think that MiniNauts is better at helping you read, compared to worksheets or flash cards?	No, I prefer using pen and paper – Yes, I love using games instead of other ways (1–5)
How much do you like to read?	I don't like reading – I love reading (1–5)
How many hours do you read by yourself, or with your family per day?	Hours (1–8)
How much do you like to play video games?	I don't like them at all – I really love them (1–5)
How many hours of video games do you play per day on average?	Hours (1–8)

Pick 3 genres (types of games) that you are most interested in	Racing
	Sports
	Puzzle/Building
	Action
	Adventure
	Platformer (eg. Mario or Rayman)
	Social
	MMO
On what devices do you play the most games?	Console (PlayStation or XBox)
	PC or Mac
	Smartphone
	Handheld (Nintendo DS or PS Vita)
	Browser Games (eg. game websites)
	Other

IV. Pre-Study Teacher Interview Questions

This appendix provides a summarised form of a teacher interview conducted between the researchers and two Primary School teachers. Key points from the interviews were transcribed and summarised.

In general, what is your process for teaching reading, reading comprehension, and inference skills? How does this process work over the course of a week, term and the wider year?

More specifically, could you describe this process in a step by step manner, with as much detail as possible?

You mentioned previously that demonstration, worked examples and repetition were important stages in your teaching process. Could you describe these stages in a step-by-step manner? What are the benefits and the downsides to this repetition?

In what ways do you reduce your teaching support over time so that students can begin to control their own learning? In what ways does this affect your students? Do all students in your classroom respond differently to this style?

How does feedback on each student's performance fit into this process? When, where and how is this feedback delivered to each student?

What are some of the challenges for teaching in this way? In what ways do the students individually respond to this teaching style?

Student's Perspective

The following questions are all about gaining an insight into what your students really feel about video games. The questions are just a guide, and you can ask the

students these questions in any format or order you think would work in your classroom.

How often do your students play video games? Are video games played more than once a day, in different contexts? For example, at home on a computer, on the bus on a phone, at home on a games console?

How long do your students play video games for in one session?

What would your students like from a video game that teaches reading comprehension? Would they like to learn, have fun, and spend time doing something different?

Do your students like to master one thing in a video game, or do they enjoy a lot of variety?

What kind of video games do your students like to play at home and at school? Do they like action, puzzles, stories, racing etc.?

What parts of these video games do the students enjoy? Do they like having fun, getting lost in another world, competing with people, getting the best score?

Do your students like games about a hero? Do your students enjoy games set in different worlds or time periods?

Do your students like to know how well they are doing in the game? Do they like high-scores? Do they like the game telling them they've done well? How about if the game tell tells them they did something wrong; do they like that?

What parts of video games do your students dislike? What bits of video games would they now want to see in this new game?

If your students played a game that was focused on reading comprehension and inference, what stories (or what kind of stories) would they like to read about? Would they prefer fiction or non-fiction?

V. Post-Study Teacher Interview Questions

7. Did you find that the Mini-Game System (MGS) was overall an effective tool in your classroom?
    a. If so, what was the most effective feature or interaction?
    b. If not, what caused the system to be ineffective in your opinion?
    c. In your opinion, did the MGS suit males as well as females?
8. Do you believe that the students who used the MGS received benefits to their reading comprehension skillset?
    a. If so, what benefits were they?
    b. If not, why do you think students did not benefit?
    c. Was this benefit shared amongst all students
        i. If not, why do you think this is the case?
9. In your opinion, did students personally enjoy using the MGS overall?
    a. If so, what aspects do you believe the students enjoyed the most?
    b. If not, what feature or interaction would you identify as being the cause of this lack of enjoyment?
10. Did the students enjoy the gameplay experience?
    a. Was the enjoyment shared homogenously, or did certain demographics enjoy the game more/less?
    b. Did the length of each Mini-Game suit the needs of the classroom activity plan?
11. How effective do you believe the exposure time was for students using this system?
    a. Do you believe that students should be exposed to this system more or less often, and if so, why?
12. If you could perform this intervention again, what aspects would you change, to gain a greater benefit from your students?
    a. Why would you make these changes?
    b. Would you recommend the MGS be expanded to include other topics areas, and more mini-games?

Bibliography

Abrams, S. & Gerber, H., 2013. Achieving through the Feedback Loop: Videogames, Authentic Assessment, and Meaningful Learning.

Abrams, S.S., Chen, X.J. & Downton, M.P., 2018. *Managing Educational Technology: School Partnerships and Technology Integration*, Routledge.

ABS, 2016a. 2016 Census QuickStats. Available at: http://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/6?opendocument [Accessed 28–2, 2019].

ABS, 2016b. Household Use of Information Technology, Australia, 2016-17. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 [Accessed 5–10, 2018].

ACARA, 2017a. About Naplan. Available at: https://www.nap.edu.au/about [Accessed 3–11, 2017].

ACARA, 2017b. Australian Curriculum., pp.1–180.

ACARA, 2018. MySchool Rankings. Available at: https://www.myschool.edu.au/ [Accessed 5–10, 2018].

ACARA, 2014. Naplan 2014 Results. Available at: http://reports.acara.edu.au/Home/Results.

ACARA, 2017c. NHMRC Statement on Data Sharing. Available at: https://www.nhmrc.gov.au/grants-funding/policy/nhmrc-statement-data-sharing [Accessed 3, 2017].

ACARA, 2017d. Organising Elements of Literacy. Available at: http://v7-5.australiancurriculum.edu.au/generalcapabilities/literacy/organising-elements/organising-elements [Accessed 6–11, 2017].

ACARA, 2017e. Statements of Learning for English. Available at: http://www.curriculum.edu.au/verve/_resources/SOL_English_Copyright_update2008_file.pdf [Accessed 14–3, 2017].

ACER, Progressive Achievement Tests in Reading. Available at: https://www.acer.org/pat/tests/reading [Accessed 14–10, 2017].

Adams, P.C., 1998. Teaching and learning with SimCity 2000. *Journal of Geography*, 97(2), pp.47–55.

Aleven, V. & Koedinger, K.R., 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive science*,

26(2), pp.147–179.

Alfassi, M., 1998. Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading classes. *American Educational Research Journal*, 35(2), pp.309–332.

Amory, A., 2007. Game object model version II: a theoretical framework for educational game development. *Educational Technology Research and Development*, 55(1), pp.51–77.

Amory, A. & Naicker, K., 1999. The use of computer games as an educational tool: identification of appropriate game types and game elements. *… Journal of Educational …*, 30(4), pp.311–321.

Anderson, C.G. et al., 2018. Failing up: How failure in a game environment promotes learning through discourse. *Thinking Skills and Creativity*.

Anderson, R.C. & Pearson, P.D., 1984. A schema-theoretic view of basic processes in reading comprehension. *Handbook of reading research*, 1, pp.255–291.

Anderson, R.C., Wilson, P.T. & Fielding, L.G., 1988. Growth in reading and how children spend their time outside of school. *Reading research quarterly*, pp.285–303.

Annema, J.-H. et al., 2012. Design of web-based mini-games for language learning: an evidence-based and user-centred approach. In *Serious Games: The Challenge*. Springer, pp. 79–81.

Annetta, L.A., 2010. The "I's" have it: A framework for serious educational game design. *Review of General Psychology*, 14(2), p.105.

Anolli, L. et al., 2010. Emotions in serious games: From experience to assessment. *International Journal of Emerging Technologies in Learning (iJET)*, 5(2010).

Applegate, A.J. & Applegate, M.D., 2010. A study of thoughtful literacy and the motivation to read. *The Reading Teacher*, 64(4), pp.226–234.

Arnab, S. et al., 2015. Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2), pp.391–411.

Arnab, S. et al., 2013. The development approach of a pedagogically-driven serious game to support Relationship and Sex Education (RSE) within a classroom setting. *Computers & Education*, 69, pp.15–30.

Aronson, J., 1995. A pragmatic view of thematic analysis. *The qualitative report*, 2(1), pp.1–3.

Australia, E.S., 2017. Improve. Available at: http://www.improve.edu.au/ [Accessed 3–11, 2017].

Australian Bureau of Statistics, 2012. Adult Literacy and Life Skills Survey, Summary Results, Australia, 2006. Available at: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4228.0Main+Features12006%20(Reissue)?OpenDocument [Accessed 10–2, 2015].

Australian Bureau of Statistics, 2010. Tasmanian State and Regional Indicators, Dec 2010. Available at: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1307.6Main+Features9Dec+2010.

Backlund, P. & Hendrix, M., 2013. Educational games-are they worth the effort? A literature survey of the effectiveness of serious games. In *Games and virtual worlds for serious applications (VS-GAMES), 2013 5th international conference on*. pp. 1–8.

Baek, Y.K., 2008. What hinders teachers in using computer and video games in the classroom? Exploring factors inhibiting the uptake of computer and video games. *CyberPsychology & Behavior*, 11(6), pp.665–671.

Bakar, N.A. & Nosratirad, E., 2013. Sustaining vocabulary acquisition through computer game: A case study. *Asian Social Science*, 9(5), p.235.

Bakker, M., 2014. *Using mini-games for learning multiplication and division: a longitudinal effect study*.

Baranowski, T. et al., 2008. Playing for real: video games and stories for health-related behavior change. *American journal of preventive medicine*, 34(1), pp.74–82.

Beavis, C., 2017. Serious Play: Literacy, Learning and Digital Games. In *Serious Play*. Routledge, pp. 17–34.

Becker, K., 2007. Digital game-based learning once removed: Teaching teachers. *British Journal of Educational Technology*, 38(3), pp.478–488.

Biesta, G. & Burbules, N.C., 2003. *Pragmatism and educational research*, Rowman \& Littlefield Lanham, MD.

Biggs, J. & Tang, C., 2011. *Teaching For Quality Learning At University*, McGraw-Hill Education. Available at: http://books.google.com.au/books?id=XhjRBrDAESkC.

Black, P. & Wiliam, D., 1998. *Inside the black box: Raising standards through classroom assessment*, Granada Learning.

Blake eLearning, 2017. Reading Eggs. Available at: https://readingeggs.com.au/ [Accessed 14–2, 2016].

Bloom, B.S. et al., 1956. *Taxonomy of educational objectives, handbook I: The cognitive domain*, New York: David McKay Co Inc.

Blumberg, F.C. et al., 2013. Serious games: What are they? What do they do? Why should we play them. *Oxford handbook of media psychology. Oxford University Press, New York*, pp.334–351.

Breuer, J. & Bente, G., 2010. Why so serious? On the relation of serious games and learning. *Journal for Computer Game Culture*, 4, pp.7–24.

Calder, B.J., Phillips, L.W. & Tybout, A.M., 1982. The concept of external validity. *Journal of Consumer Research*, 9(3), pp.240–244.

Callaghan, M.J. et al., 2016. Practical application of the Learning Mechanics-Game Mechanics (LM-GM) framework for Serious Games analysis in engineering education. In *Remote Engineering and Virtual Instrumentation (REV), 2016 13th International Conference on*. pp. 391–395.

Campbell, A. et al., 1996. *Assessing Literacy: The Framework for the National Adult Literacy Survey*, Diane Publishing Company. Available at: https://books.google.com.au/books?id=5_9Hk8kP46UC.

Cannon-Bowers, J., 2010. *Serious Game Design and Development: Technologies for Training and Learning: Technologies for Training and Learning*, IGI Global.

Chaiklin, S., 2003. The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's educational theory in cultural context*, 1, pp.39–64.

Charsky, D., 2010. From edutainment to serious games: A change in the use of game characteristics. *Games and Culture*.

Charsky, D. & Mims, C., 2008. Integrating commercial off-the-shelf video games into school curriculums. *TechTrends*, 52(5), pp.38–44.

Charsky, D.G., 2004. Evaluation of the effectiveness of integrating concept maps and computer games to teach historical understanding.

Chen, S. & Michael, D., 2005. Proof of learning: Assessment in serious games. *Retrieved October*, 17, p.2008.

Chou, C. & Tsai, M.-J., 2007. Gender differences in Taiwan high school students' computer game playing. *Computers in Human Behavior*, 23(1), pp.812–824.

Cockburn, A., 2002. *Agile software development*, Addison-Wesley Boston.

Cohen, P.A., 1980. Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in higher education*, 13(4), pp.321–341.

Coltheart, M. & Prior, M., 2006. Learning to read in Australia. *Australian Journal of Learning Difficulties*, 11(4), pp.157–164.

Connolly, B.E.M., 2012a. A systematic literature review of empirical evidence on computer games and serious games. *Computers \& Education*, 59(2), pp.661–686.

Connolly, B.E.M., 2012b. A systematic literature review of empirical evidence on computer games and serious games. *Computers \& Education*, 59(2), pp.661–686. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0360131512000619.

Cowley, B., Heikura, T. & Ravaja, N., 2013. Learning loops-interactions between guided reflection and experience-based learning in a serious game activity. *Journal of Computer Assisted Learning*, 29(4), pp.348–370.

Cox, K.E. & Guthrie, J.T., 2001. Motivational and cognitive contributions to students' amount of reading. *Contemporary Educational Psychology*, 26(1), pp.116–131.

Cree, A., Kay, A. & Steward, J., 2012. The Economic and Social Cost of Illiteracy: A Snapshot of Illiteracy in a Global Context.

Creswell, J.W. et al., 2003. Advanced mixed methods research designs. *Handbook of mixed methods in social and behavioral research*, 209, p.240.

Creswell, J.W., 2013. *Research design: Qualitative, quantitative, and mixed methods approaches*, Sage publications.

Creswell, J.W. & Clark, V.L.P., 2007. Designing and conducting mixed methods research.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience Harper \& Row New York Google Scholar.

Dede, C., 2005. Planning for neomillennial learning styles: Implications for investments in technology and faculty. In Oblinger, Diana G. and Oblinger, James L. and Windham, Carie, ed. *Educating the net generation*. Boulder and CO: EDUCAUSE, pp. 15.1–15.22.

Department of Education, 2011. An Evaluation of Digital Game-Based Learning in school contexts. Available at: http://clic.det.nsw.edu.au/clic/documents/managahigh_exec_sum.pdf [Accessed 6–1, 2015].

Department of Education, 2017. Guidelines for Conduction Research in Tasmanian Government Educational Settings. Available at: https://documentcentre.education.tas.gov.au/Documents/Guidelines-for-Conducting-Research.pdf [Accessed 2–8, 2017].

Deterding, S. et al., 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. pp. 9–15.

Devisch, O. et al., 2018. Mini is beautiful: Playing serious mini-games to facilitate collective learning on complex urban processes.

Dimitrov, D.M. & Rumrill Jr, P.D., 2003. Pretest-posttest designs and measurement of change. *Work*, 20(2), pp.159–165.

Dominguez, A. et al., 2013. Gamifying learning experiences: Practical implications and outcomes. *Computers \& Education*, 63, pp.380–392.

Dormans, J. & others, 2012. *Engineering emergence: applied theory for game design*, Universiteit van Amsterdam [Host].

Doyle, L., Brady, A.-M. & Byrne, G., 2009. An overview of mixed methods research. *Journal of Research in Nursing*, 14(2), pp.175–185.

Duffy, T.M. & Jonassen, D.H., 2013. *Constructivism and the technology of instruction: A conversation*, Routledge.

Eagle, M. & Barnes, T., 2009. Experimental evaluation of an educational game for improved learning in introductory computing. In *ACM SIGCSE Bulletin*. pp. 321–325.

Edwards, R., Hanson, A. & Raggatt, P., 2013. *Boundaries of Adult Learning*, Taylor \& Francis. Available at: http://books.google.com.au/books?id=MVvdAAAAQBAJ.

Ertmer, P.A. et al., 2012. Teacher beliefs and technology integration practices: A critical relationship. *Computers \& Education*, 59(2), pp.423–435.

Euston, J., 2014. Gaming: The Lingua Franca of Mobile. Available at: http://www.flurry.com/blog/flurry-insights/gaming-lingua-franca-mobile#.VBxYzC5dVV6.

Feilzer, M.Y., 2010. Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of mixed methods research*, 4(1), pp.6–16.

Forum, I. et al., 2005. Educational Technology \& Society., 8(2).

Frazer, A., Argles, D. & Wills, G., 2007a. Assessing The Usefulness Of Mini-games As Educational Resources. Available at: http://eprints.ecs.soton.ac.uk/14174.

Frazer, A., Argles, D. & Wills, G., 2007b. Is less actually more? The usefulness of educational mini-games. In *Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on*. pp. 533–537.

De Freitas, S. & Jarvis, S., 2006. A framework for developing serious games to meet learner needs.

De Freitas, S. & Liarokapis, F., 2011. Serious games: a new paradigm for education? In *Serious games and edutainment applications*. Springer, pp. 9–23.

Gambrell, L.B., 2011. Seven rules of engagement: What's most important to know about motivation to read. *The reading teacher*, 65(3), pp.172–178.

Game Ontology Project, 2017. Game Ontology Project. Available at: http://www.gameontology.com/index.php/Main_Page [Accessed 12–9, 2017].

Garris, R., Ahlers, R. & Driskell, J.E., 2002. Games, Motivation, and Learning: A Research and Practice Model. *Simulation \& Gaming*, 33(4), pp.441–467. Available at: http://sag.sagepub.com/cgi/doi/10.1177/1046878102238607.

Garris, R. & Driskell, 2002. Games, motivation, and learning: A research and practice model. *Simulation \& gaming*, 33(4), pp.441–467.

Garrison, C. & Ehringhaus, M., 2007. Formative and summative assessments in the classroom.

Gee, J.P., 2005. Good video games and good learning. In *Phi Kappa Phi Forum*. p. 33.

Girard, C., Ecalle, J. & Magnan, A., 2013. Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3), pp.207–219.

Goguen, J.A. & Linde, C., 1993. Techniques for requirements elicitation. In *[1993] Proceedings of the IEEE International Symposium on Requirements Engineering*. pp. 152–164.

Greenberg, B.S. et al., 2010. Orientations to video games among gender and age groups. *Simulation & Gaming*, 41(2), pp.238–259.

Grimshaw, J. et al., 2000. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family practice*, 17(90001), pp.11–16.

Halai, A., 2006. Ethics in qualitative research: Issues and challenges. *EdQual A Research Programme Consortium on Implementing Education Quality in Low Income Countries. EdQual Working Paper Number*, 4, pp.1–12.

Hamari, J., Koivisto, J. & Sarsa, H., 2014. Does gamification work?-a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*. pp. 3025–3034.

Hartmann, T. & Klimmt, C., 2006. Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication*, 11(4), pp.910–931.

Hattie, J., 1999. Influences on student learning. *Inaugural lecture given on August*, 2, p.1999.

Hattie, J. & Timperley, H., 2007. The power of feedback. *Review of educational research*, 77(1), pp.81–112.

Health, N. & Council, M.R., 2017. Children and Young People.

Hébert, C. & Jenson, J., 2017. Digital Game-Based Pedagogy: Exploring Teaching Strategies for Classroom Teachers in the use of Video Games in K-12 Classrooms. In *ECGBL 2017 11th European Conference on Game-Based Learning*. p. 227.

Henry, K., Lagos, A. & Berndt, F., 2012. Bridging the literacy gap between boys and girls: An opportunity for the National Year of Reading 2012. *Australian Library Journal*, 61(2), pp.143 – 150. Available at: https://login.ezproxy.utas.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=azh&AN=76292628&site=eds-live.

Hess, T. & Gunter, G., 2013. Serious game-based and nongame-based online courses: Learning experiences and outcomes. *British Journal of Educational Technology*, 44(3), pp.372–385.

Huang, W.-H., Huang, W.-Y. & Tschopp, J., 2010. Sustaining iterative game playing processes in DGBL: The relationship between motivational processing and outcome processing. *Computers & Education*, 55(2), pp.789–797. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0360131510000977.

Huizenga, J. et al., 2009. Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game. *Journal of Computer Assisted Learning*, 25(4), pp.332–344.

Hunter, H., 2013. *Shift: Using Business Simulations and Serious Games: A straightforward guide for corporate training and development*, Hugh Hunter.

Hwang, G.-J., Yang, L.-H. & Wang, S.-Y., 2013. A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, 69, pp.121–130.

Hwong, C., 2016. The Average Mobile Game Day. Available at: http://www.vertoanalytics.com/average-mobile-game-day/ [Accessed 22–10, 2017].

Ibáñez, M.-B., Di-Serio, A. & Delgado-Kloos, C., 2014. Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on Learning Technologies*, 7(3), pp.291–301.

IJsselsteijn, W., De Kort, Y. & Poels, K., 2013. The game experience questionnaire: Development of a self-report measure to assess the psychological impact of digital games. manuscript in preparation.

Illanas Vila, A.I. et al., 2008. Conceptual mini-games for learning.

Imbellone, A., Botte, B. & Medaglia, C.M., 2015. Serious games for mobile devices: the intouch project case study. *International Journal of Serious Games*, 2(1), pp.17–

27.

Isabel, A. et al., 2014. Un videojuego como sistema predictivo de aprendizaje.

Ivankova, N.V., Creswell, J.W. & Stick, S.L., 2006. Using mixed-methods sequential explanatory design: From theory to practice. *Field methods*, 18(1), pp.3–20.

Ivey, G. & Broaddus, K., 2001. "Just Plain Reading": A Survey of What Makes Students Want to Read in Middle School Classrooms. *Reading Research Quarterly*, 36(4), pp.350–377. Available at: http://www.jstor.org/stable/748056.

De Jans, S. et al., 2017. Using games to raise awareness: How to co-design serious mini-games? *Computers \& Education*, 110, pp.77–87.

Jarvin, L., 2015. Edutainment, games, and the future of education in a digital world. *New directions for child and adolescent development*, 2015(147), pp.33–40.

Jewitt, C. et al., 2001. Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review*, 53(1), pp.5–18.

Jewitt, C., Bezemer, J.J. & O'Halloran, K.L., 2016. *Introducing multimodality*, Routledge London.

Johnson, R.B. & Onwuegbuzie, A.J., 2004. Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), pp.14–26.

Kapp, K.M., 2012. *The gamification of learning and instruction: game-based methods and strategies for training and education*, John Wiley \& Sons.

Keenan, J.M., Betjemann, R.S. & Olson, R.K., 2008. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), pp.281–300.

Kervin, L., 2016. Powerful and playful literacy learning with digital technologies. *Australian Journal of Language and Literacy*, (1), p.64. Available at: https://login.ezproxy.utas.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.445983114&site=eds-live.

Kiili, K., 2005. Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education*, 8(1), pp.13–24.

Kiili, K., 2007. Foundation for problem-based gaming. *British Journal of Educational Technology*, 38(3), pp.394–404. Available at: http://doi.wiley.com/10.1111/j.1467-8535.2007.00704.x http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.2007.00704.x/full.

Kiili, K. & Lainema, T., 2008. Foundation for measuring engagement in educational games. *Journal of Interactive Learning Research*, 19(3), pp.469–488.

Kintsch, W., 1994. Text comprehension, memory, and learning. *American Psychologist*, 49(4), p.294.

Kluger, A.N. & DeNisi, A., 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2), p.254.

Kolb, D.A. et. al, 1984. *Experiential learning: Experience as the source of learning and development*, Prentice-Hall Englewood Cliffs, NJ.

Kong, K. et al., 2010. Gee's Learning Principles for Good Games: An Analysis of how Gee's Learning Principles meets Gardner's Theory of Multiple Intelligences. In *TCC-Teaching Colleges and Community Worldwide Online Conference*. pp. 159–167.

Koops, M. & Hoevenaar, M., 2012. Conceptual change during a serious game: Using a Lemniscate model to compare strategies in a physics game. *Simulation \& Gaming*, p.1046878112459261.

Korozi, M. et al., 2012. Ambient educational mini-games. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*. p. 802.

Landers, R.N. & Callan, R.C., 2011. Casual social games as serious games: The psychology of gamification in undergraduate education and employee training. In *Serious games and edutainment applications*. Springer, pp. 399–423.

Lê, T. & Lê, Q., 2007. Using computers to promote literacy development. *technology and teaching*.

Learning, 3P, 2017. Mathletics. Available at: http://au.mathletics.com/ [Accessed 23–4, 2017].

Lieberman, M., 2010. Four ways to teach with video games. *Retrieved March*, 25, p.2014.

Lim, T. et al., 2016. The lm-gm framework for serious games analysis.

Lucas, K. & Sherry, J.L., 2004. Sex differences in video game play: A communication-based explanation. *Communication research*, 31(5), pp.499–523.

Ma, M., Oikonomou, A. & Jain, L.C., 2011. *Serious games and edutainment applications, Second Edition*, Springer.

MacArthur, C.A. et al., 2001. Technology Applications for Students with Literacy Problems: A Critical Review. *The Elementary School Journal*, 101(3), pp.pp. 273–301. Available at: http://www.jstor.org/stable/1002248.

Machet, M., 2002. Addressing problems of literacy in disadvantaged communities. *Language Matters* , 33(1), pp.1–24.

Marchand, A. & Hennig-Thurau, T., 2013. Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities. *Journal of Interactive Marketing*, 27(3), pp.141–157.

Marfisi-Schottman, I., George, S. & Tarpin-Bernard, F., 2010. Tools and methods for efficiently designing serious games. In *Proceedings of the 4th Europeen Conference on Games Based Learning ECGBL*. pp. 226–234.

Marsh, J. & Hallet, E., 2008. *Desirable literacies: Approaches to language and literacy in the early years*, Sage.

Marsh, T., 2011. Serious games continuum: Between games for purpose and experiential environments for purpose. *Entertainment Computing*, 2(2), pp.61–68. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1875952110000224.

Mathworld, W.A., 2017. Bonferonni Correction. Available at: http://mathworld.wolfram.com/BonferroniCorrection.html [Accessed 9–6, 2016].

Maughan, B. et al., 2009. Persistence of literacy problems: spelling in adolescence and at mid-life. *Journal of Child Psychology and Psychiatry*, 50(8), pp.893–901.

Mayer, I. et al., 2014. The research and evaluation of serious games: Toward a comprehensive methodology. *British Journal of Educational Technology*, 45(3), pp.502–527.

Mayer, I., 2012. Towards a comprehensive methodology for the research and evaluation of serious games. *Procedia Computer Science*, 15, pp.233–247.

McCutchen, D. et al., 2002. Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of learning disabilities*, 35(1), pp.69–86.

McKenna, M.C., 2006. *International Handbook of Literacy and Technology: Volume two*, Lawrence Erlbaum Associates. Available at: https://books.google.com.au/books?id=84RWuXAINAgC.

McMillan, J. & Marks, G.N., 2003. *School Leavers in Australia: Profiles and Pathways. LSAY Research Report.*, ERIC.

Media, C.S., 2014. Children, Teens, and Reading. Available at: https://www.commonsensemedia.org/file/csm-childrenteensandreading-2014pdf/download [Accessed 10, 2017].

Mekler, E.D. et al., 2015. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*.

Meyer, B., 2009. Learning English through serious games-reflections on teacher and learner performance. In *Transactions on Edutainment III*. Springer, pp. 82–92.

Mezirow, J. et al., 1990. How critical reflection triggers transformative learning. *Fostering critical reflection in adulthood*, pp.1–20.

Michael, D.R. & Chen, S.L., 2005. *Serious games: Games that educate, train, and inform*, Muska & Lipman/Premier-Trade.

Michie, S. et al., 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine*, 46(1), pp.81–95.

Michie, S., Atkins, L. & West, R., 2014. The behaviour change wheel: a guide to designing interventions. *Needed: physician leaders*, 26.

Michie, S., van Stralen, M.M. & West, R., 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, 6(1), p.42.

Modern Dream, 2013. The Typing of the Dead: Overkill. Available at: http://store.steampowered.com/app/246580/The_Typing_of_The_Dead_Overkill/.

Moreno, R. & Mayer, R.E., 2005. Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *Journal of educational psychology*, 97(1), p.117.

Mozelius, P., 2014. What can be Learned From Playing Digital Games Outside School? In *European Conference on Games Based Learning*. p. 415.

Murphy, C., 2011. Why games work and the science of learning. In *Interservice, Interagency Training, Simulations, and Education Conference*.

Nectar Directorate, 2017. Nectar Research Cloud. Available at: https://nectar.org.au/about/ [Accessed 2–8, 2016].

Nevin, J.A. & Grace, R.C., 2000. Behavioral momentum: Empirical, theoretical, and metaphorical issues. *Behavioral and Brain Sciences*, 23(1), pp.117–125.

Nicholson, S., 2012. A user-centered theoretical framework for meaningful gamification. *Games, Learning, Society*, 8(1), pp.223–230.

Okan, Z., 2003. Edutainment: is learning at risk? *British Journal of Educational Technology*, 34(3), pp.255–264.

Pagulayan, R.J. et al., 2002. *User-centered design in games*, CRC Press Boca Raton, FL.

Panagiotakopoulos, C.T., 2011. Applying a conceptual mini game for supporting simple mathematical calculation skills: Students' perceptions and considerations. *World Journal of Education*, 1(1), p.p3.

Papastergiou, M., 2009. Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers \& Education*, 52(1), pp.1–12.

Paras, B., 2005. Game, motivation, and effective learning: An integrated model for educational game design. Available at: http://summit.sfu.ca/item/281.

Parkin, S., 2017. How video game difficulty became a cultural battleground. Available at: http://www.eurogamer.net/articles/2017-10-21-how-video-game-difficulty-became-a-cultural-battleground [Accessed 4–10, 2017].

Pavlas, D., 2010. A model of flow and play in game-based learning: The impact of game characteristics, player traits, and player states.

Perez, S., 2014. Time Spent In Apps Up 21% Over Last Year. Available at: https://techcrunch.com/2014/09/16/time-spent-in-apps-up-21-over-last-year/ [Accessed 18–4, 2016].

Perfetti, C. & Adlof, S.M., 2012. Reading comprehension: A conceptual framework from word meaning to text meaning. *Measuring up: Advances in how to assess reading ability*, pp.3–20.

Perfetti, C.A. & Marron, M.A., 1995. Learning To Read: Literacy Acquisition by Children and Adults.

Poels, Y. & Annema, J., 2012. Developing User-Centered Video Game Concepts for Language Learning. *ACHI 2012, The Fifth \ldots*, (c), pp.11–16. Available at: http://www.thinkmind.org/index.php?view=article\&articleid=achi_2012_1_30_20253.

Prensky, M., 2005. Complexity matters. *Educational Technology*, 45(4), pp.5–20.

Prensky, M., 2001. *Digital Game-based Learning*, McGraw-Hill. Available at: http://books.google.com/books?id=XBwiAQAAIAAJ.

Prensky, M., 2008. Students as designers and creators of educational computer games: Who else? *British Journal of Educational Technology*, 39(6), pp.1004–1019.

Purcell-Gates, V. et al., 2002. Impact of authentic adult literacy instruction on adult literacy practices. *Reading Research Quarterly*, 37(1), pp.70–92.

Rapp, A., 2017. Drawing inspiration from World of Warcraft: Gamification design elements for behavior change technologies. *Interacting with Computers*, pp.1–31.

Ray, B. & Coulter, G.A., 2010. Perceptions of the value of digital mini-games: Implications for middle school classrooms. *Journal of Computing in Teacher Education*, 26(3), pp.92–100.

readable.io, 2017. Readable.io. Available at: https://readable.io/text/ [Accessed 3–11, 2017].

Resnick, M., 2004. Edutainment? No thanks. I prefer playful learning. *Associazione Civita Report on Edutainment*, 14, pp.1–4.

Richmond, M., Robinson, C. & Sachs-Israel, M., 2012. The global literacy challenge.

Rieber, L.P., 1996. Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational technology research and development*, 44(2), pp.43–58.

Rivera-Batiz, F.L., 1992. Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, pp.313–328.

Roueche, J.E. & Roueche, S.D., 1999. *High stakes, high performance: Making remedial education work.*, ERIC.

Rupp, A.A., Ferne, T. & Choi, H., 2006. How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language testing*, 23(4), pp.441–474.

Ryan, R.M., Rigby, C.S. & Przybylski, A., 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, 30(4), pp.344–360.

Sadler, D.R., 1989. Formative assessment and the design of instructional systems. *Instructional science*, 18(2), pp.119–144.

Sarroub, L.K., Pearson, P.D. & others, 1998. Reading Comprehension Assessment.

Seale, J.K. & Cann, A.J., 2000. Reflection on-line or off-line: the role of learning technologies in encouraging students to reflect. *Computers \& Education*, 34(3), pp.309–320.

Segers, E. & Verhoeven, L., 2002. Multimedia support of early literacy learning. *Computers & Education*, 39(3), pp.207–221.

Serafini, F., 2012. Reading multimodal texts in the 21st century. *Research in the Schools*, 19(1), pp.26–32.

Shaffer, D.W., 2006. Epistemic frames for epistemic games. *Computers \& education*, 46(3), pp.223–234.

Sharp, S. & Fitz-Gerald, M., 2016. Grammatikus.

Snow, C.E. & Matthews, T.J., 2016. Reading and Language in the Early Grades Reading and Language in the Early Grades. *Future of Children*, 26(1), pp.57 – 74. Available at: https://eric.ed.gov/?id=EJ1118540.

Sorapure, M., Inglesby, P. & Yatchisin, G., 1998. Web literacy: Challenges and opportunities for research in a new medium. *Computers and Composition*, 15(3), pp.409–424.

Spörer, N., Brunstein, J.C. & Kieschke, U., 2009. Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction*, 19(3), pp.272–286.

Sqire, K., 2005. Changing the game: What happens when video games enter the classroom? *Innovate: Journal of online education*, 1(6), p.5.

Steinkuehler, C., Compton-Lilly, C. & King, E., 2010. Reading in the context of online games. In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1*. p. 228.

Steinkuehler, C. & Squire, K., 2014. Videogames and learning. *Cambridge handbook of the learning sciences*, pp.377–396.

Stiggins, R.J., 2002. Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), pp.758–765.

Susi, T., Johannesson, M. & Backlund, P., 2007. Serious games: An overview.

Sutton-Smith, B., 2009. *The ambiguity of play*, Harvard University Press.

Tjus, T., Heimann, M. & Nelson, K.E., 1998. Gains in Literacy through the Use of a Spcially Developed Multimedia Computer Strategy Positive Findings from 13 Children with Autism. *Autism*, 2(2), pp.139–156.

Topping, K.J., 2015. Fiction and Non-Fiction Reading and Comprehension in Preferred Books. *Reading Psychology*, 36(4), pp.350 – 387. Available at: http://www.tandfonline.com/doi/abs/10.1080/02702711.2013.865692.

Trigwell, K., Prosser, M. & Waterhouse, F., 1999. Relations between teachers' approaches to teaching and students' approaches to learning. *Higher education*, 37(1), pp.57–70.

Trochim, W., 2006. The T-Test. Available at: https://www.socialresearchmethods.net/kb/stat_t.php [Accessed 1–12, 2016].

Watson, W. & Yang, S., 2016. Games in schools: Teachers' perceptions of barriers to game-based learning. *Journal of Interactive Learning Research*, 27(2), pp.153–170.

Weaver, C., 1988. *Reading process and practice: From socio-psycholinguistics to whole language.*, ERIC.

Weis, M., Heikamp, T. & Trommsdorff, G., 2013. Gender differences in school achievement: The role of self-regulation. *Frontiers in Psychology*, 4.

Westera, W. et al., 2008. Serious games for higher education: a framework for reducing design complexity. *Journal of Computer Assisted Learning*, 24(5), pp.420–432.

White, B.Y. & Frederiksen, J.R., 1998. Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and instruction*, 16(1), pp.3–118.

Willenberg, I., 2005. Starting at the beginning: Early childhood literacy intervention as a strategy for reducing adult illiteracy. *Language Matters: Studies in the Languages of Southern Africa*, 36(2), pp.162–175.

Williams, E., 2014. Breaking the Barriers to Reading Success in Middle and High Schools. *Reading Improvement*, 51(2), pp.233 – 236. Available at: https://eric.ed.gov/?id=EJ1034942.

Woessner, M., 2015. Teaching with SimCity: Using sophisticated gaming simulations to teach concepts in introductory American Government. *PS: Political Science \& Politics*, 48(2), pp.358–363.

Wood, L.E., 1997. Semi-structured interviewing for user-centered design. *interactions*, 4(2), pp.48–61.

Wouters, P. et al., 2013. A meta-analysis of the cognitive and motivational effects of serious games.

Yuill, N. & Oakhill, J., 1988. Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, 2(1), pp.33–45.

Zagal, J.P. & Bruckman, A., 2008. The game ontology project: Supporting learning while contributing authentically to game studies. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 2*. pp. 499–506.

Zapata-Rivera, J.-D. & Greer, J., 2003. Analyzing student reflection in the learning game. In *S. Bull, P. Brna \& V. Dimitrova (Eds.) Proceedings of the AIED'03-Workshop on Learner Modelling for Reflection*. pp. 288–298.

Zebel, S., Ufkes, E. & Scheepens, R., 2013. Testing and Feedback in DGL. *essay.utwente.nl*. Available at: http://essay.utwente.nl/64618/1/Saman%20Yan%20D.%20-%20S1189735%20-%20masterscriptie.pdf.